numero 224
aprile 1987
anno xx
volume xxxviii

La fame negli Stati Uniti

Pressoché eliminata negli anni settanta, si ripresenta oggi come una conseguenza delle restrizioni federali; secondo la comune definizione di fame, ne soffrono circa 12 milioni di bambini e otto milioni di adulti

di J. Larry Brown

Non c'è una definizione universalmente valida di fame, ma un concetto generalmente accettato in medicina è quello secondo il quale un individuo che soffra la fame ha una carenza cronica di quelle sostanze nutritive che sono necessarie per la crescita e per vivere in buona salute. Se, in base a questa definizione, ci si chiede quanti statunitensi soffrano la fame, ci si imbatte in un dato statistico sbalorditivo: 12 milioni di bambini e otto milioni di adulti, ovvero circa il 9 per cento della popolazione. Non è da escludere che alcuni di questi individui soffrano la fame per ignoranza o per indifferenza, ma per la maggior parte di essi le ragioni vanno ricercate in un'economia che lascia molte famiglie al di sotto della soglia di povertà e in un sistema di assistenza sociale che dà loro un aiuto insufficiente.

Colpisce in modo particolare il fatto che una diffusione della fame su così vasta scala fosse già stata individuata trent'anni or sono e che le contromisure prese dal Governo degli Stati Uniti negli anni sessanta e settanta avessero in pratica eliminato il problema. Il programma di assistenza con buoni per alimenti (*food-stamp program*) fu ampliato in modo da raggiungere 20 milioni di persone. Furono istituiti programmi intesi a raggiungere e a dare da mangiare ad anziani soli. Furono incrementati i programmi di refezione scolastica per la prima e per la seconda colazione (*school lunch and breakfast programs*) così che i bambini poveri avessero la base nutritiva necessaria per un buon rendimento a scuola. Fu anche varato un programma integrativo destinato a fornire alimenti e consigli nutrizionali a madri e gestanti povere e, inoltre, cibo ai loro piccoli.

Nel 1981 l'Amministrazione Reagan incominciò a ridurre e a modificare questi programmi e già nel 1982 i segni della fame erano molto diffusi. Un numero sempre maggiore di individui si rivolgeva alle chiese e agli enti sociali e lamentava di non avere abbastanza cibo. In alcune parti degli Stati Uniti i medici incominciarono a raccogliere le lamentele di un numero sempre maggiore di pazienti che avevano fame e a osservare malattie associate alla fame - fra esse anemie, tubercolosi, crescita stentata nei bambini e osteoporosi negli adulti.

Più tardi, in quello stesso anno, arrivò il primo di numerosi rapporti che documentavano l'entità e la diffusione della fame negli Stati Uniti. Complessivamente 20 studi a livello nazionale sulla fame furono svolti tra l'ottobre del 1982 e il marzo del 1986. Dei gruppi che avevano inviato una loro relazione facevano parte l'US Conference of Mayors, il National Council of Churches, l'US Department of Agriculture e l'organizzazione di cui io sono presidente: la Physician Task Force on Hunger in America, con base presso la Harvard School of Public Health. Questi studi giungevano tutti alla conclusione che la fame era tornata a essere un serio problema nazionale.

Risultanze tipiche furono portate alla luce dalla Conference of Mayors, la conferenza dei sindaci, che verso la fine del 1984 aveva condotto un'indagine su città grandi e piccole per stabilire se e come la propagandata ripresa economica incidesse sui poveri. Quasi il 75 per cento delle amministrazioni comunali dichiarò che il bisogno di assistenza alimentare di emergenza era aumentato in quell'anno. Il numero dei centri per l'assistenza alimentare di emergenza (comprendenti le mense per i poveri, che servivano pasti, e gli spacci, che distribuivano generi alimentari) aumentò del 15 per cento nel corso dell'anno. Il numero totale dei pasti distribuiti salì del 50 per cento rispetto al 1983. Parecchie grandi città denunciarono aumenti ancora maggiori: il 200 per cento a Boston, il 182 per cento a Chicago e il 100 per cento a Dallas.

Nel novembre del 1984, il Food Research and Action Center fece un'indagine su circa 300 programmi di assistenza alimentare di emergenza e scoprì che il 65 per cento parlavano di un aumento, durante l'anno, del numero di famiglie affamate che chiedevano cibo. L'aumento del numero delle famiglie alle quali veniva dato da mangiare era superiore al 20 per cento. La maggioranza dei centri di emergenza riferì che più della metà delle persone affamate assistite era costituita da famiglie con bambini.

Nel 1985 la Physician Task Force calcolò che la fame affliggeva circa 20 milioni di americani. Pur non essendo disponibili dati più recenti, le successive testimonianze fanno pensare che il problema della fame in America si sia ulteriormente aggravato.

Gli appartenenti a determinati gruppi etnici sono particolarmente esposti ai danni della fame. È probabile che a soffrire maggiormente quando l'alimen-

Il programma di distribuzione di alimenti presso la chiesa episcopale di San Bartolomeo a New York è un esempio tipico dello sforzo che gli enti privati hanno fatto, in tutti gli Stati Uniti, per far fronte all'aumento della fame dovuto ai tagli nei programmi federali. Per cinque anni questo «programma del sabato mattina» ha distribuito a ben 275 persone pacchi viveri per integrare le loro insufficienti razioni alimentari.

Il buono per alimenti emesso dallo US Department of Agriculture rappresenta un sussidio per le famiglie bisognose in modo che possano procurarsi cibo. Deve essere consegnato ai negozi di generi alimentari per avere cibo in cambio. In seguito alle recenti modifiche apportate alle norme di assistenza, solo 19 milioni di americani ricevono questi buoni anche se sono ben 33,4 milioni coloro che vivono in povertà. Il sussidio medio per pasto è di 49 centesimi di dollaro.

tazione è inadeguata siano le gestanti, i neonati e i bambini, da una parte, e gli anziani, dall'altra. A qualsiasi età la malnutrizione può causare debolezza e letargia e può menomare il funzionamento del sistema immunitario, rendendo più vulnerabili alle malattie di natura infettiva. La mancanza di sostanze nutritive specifiche può causare una gamma di malattie da carenza che colpiscono tutti gli organi del corpo umano. Una malnutrizione grave e prolungata può portare direttamente o indirettamente alla morte.

Per una gestante la qualità e la quantità degli alimenti disponibili sono fattori determinanti ai fini di un sano sviluppo del bambino. Quando durante la gravidanza l'alimentazione è inadeguata, la madre corre numerosi pericoli. Uno di essi è l'anemia, che riflette un apporto di ferro insufficiente per le accresciute necessità della gestante. La tossiemia, un complesso di sintomi che comprende un aumento della pressione sanguigna e un rigonfiamento dei tessuti da eccessiva ritenzione di liquido, è pure un rischio grave e frequente. Recenti scoperte fanno pensare che essa sia la conseguenza di un'alimentazione inadeguata (in particolare una mancanza di proteine).

Per lo sviluppo dei tessuti e degli organi il feto ha bisogno di una dieta ricca di proteine, di vitamine e di sali minerali. Non è da escludere, quindi, che la salute del neonato possa essere compromessa da una scarsa alimentazione della madre. I rischi comprendono una nascita prematura (che ha luogo quando il neonato viene alla luce dopo 37 settimane di gestazione o prima) e il basso peso alla nascita (meno di due chilogrammi e mezzo). Nei prematuri esiste il pericolo di malattie dell'apparato respiratorio e di una debole risposta immunitaria. La malnutrizione espone anche i bambini nati a termine a problemi di salute come l'ipoglicemia (insufficienza di zucchero nel sangue), l'ipocalcemia (insufficienza di calcio nel sangue) e la policitemia (aumento dei globuli rossi) nonché, sul più lungo termine, deficienze di crescita e di sviluppo.

I pericoli ai quali è esposto un neonato gracile aumentano il rischio di morte nell'infanzia. Anche quando si fa uso di una tecnologia complessa e costosa, un bambino nato sottopeso ha 40 volte più probabilità di quelli normali di morire prima di aver compiuto un anno. Negli Stati Uniti il basso peso alla nascita è uno dei fattori che concorre per oltre la metà dei decessi che hanno luogo nel primo anno di vita e per il 75 per cento dei decessi di bambini nati da meno di un mese.

I rischi potenziali della malnutrizione per una gestante e per il feto possono essere superati fornendo alla madre, durante la gravidanza, un supplemento di alimenti. Durante la seconda guerra mondiale, quando il cibo era rigorosamente razionato, nel Regno Unito si fece in modo che le gestanti seguissero una dieta adeguata e, di fatto, la mortalità infantile registrò una flessione. In altre località, come a Leningrado e nei Paesi Bassi, dove le gestanti non ricevettero razioni supplementari particolari, il peso dei neonati diminuì e aumentò la mortalità infantile.

Studi più recenti, che analizzano gli effetti dello Special Supplemental Food Program for Women, Infants and Children (WIC), istituito dal Governo degli Stati Uniti nel 1972, dimostrano il benefico effetto sulle gestanti di una nutrizione integrata. Otto studi dello stesso genere, svolti tra il 1976 e il 1986, giunsero tutti alla conclusione che quel programma aveva avuto come risultato una salute migliore per le madri e un miglior esito delle gravidanze. Fra i risultati si ottenne una diminuzione di oltre il 20 per cento dei bambini nati sottopeso, una riduzione della mortalità fetale in gravidanza avanzata, un minor numero di nascite premature e una flessione della mortalità nel primo anno di vita.

Indipendentemente da quali siano state le condizioni della gestazione e del parto, una volta nato il bambino corre rischi notevoli se gli manca una nutrizione adeguata. Dal concepimento fino a più o meno il terzo anno di vita il cervello umano si sviluppa con particolare rapidità. In questo periodo, la funzione cerebrale può dunque venire menomata da carenze nutrizionali. Altri rischi sono la crescita stentata (definita come una statura inferiore al quinto percentile per un dato sesso e una data età) e l'atrofia (peso al di sotto del quinto percentile per età e statura). I bambini malnutriti sono, inoltre, particolarmente vulnerabili al piombo e ad altri agenti tossici presenti nell'ambiente che possono colpire il cervello e accrescere gli effetti diretti della malnutrizione sullo sviluppo intellettivo.

Come a tutte le età, la malnutrizione nella fanciullezza può indebolire la resistenza alle infezioni. I bambini scarsamente nutriti corrono il pericolo di incorrere più frequentemente in raffreddori, in infezioni dell'apparato uditivo e in altre malattie di tipo infettivo. Quindi, non è da escludere che i rischi per lo sviluppo, anche intellettuale, inerenti alla scarsa alimentazione vengano ulteriormente accresciuti dalle conseguenti assenze da scuola e dalla riduzione di alcune attività infantili.

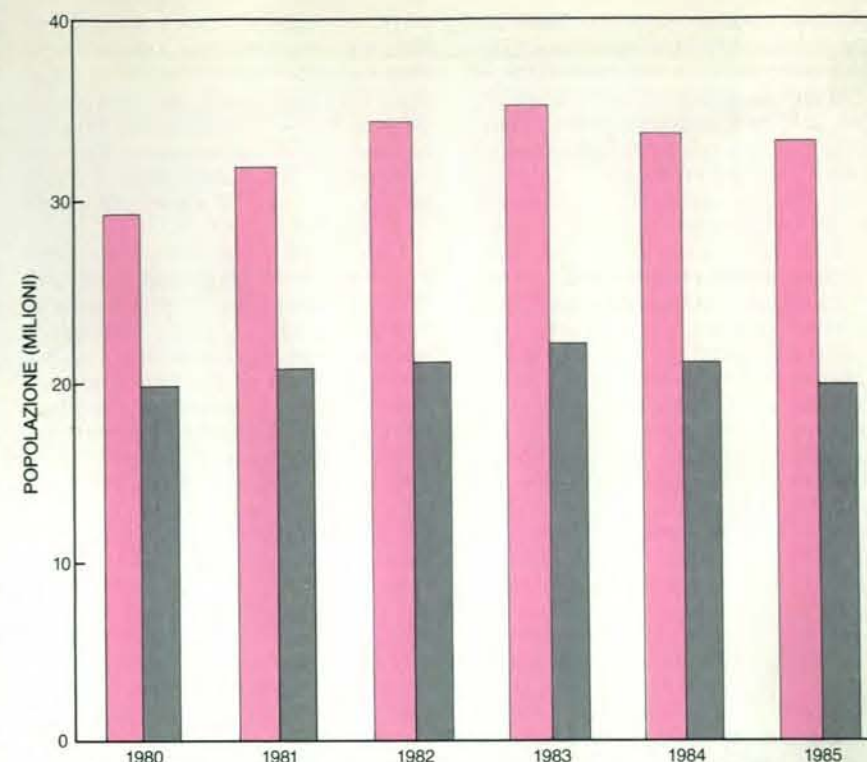
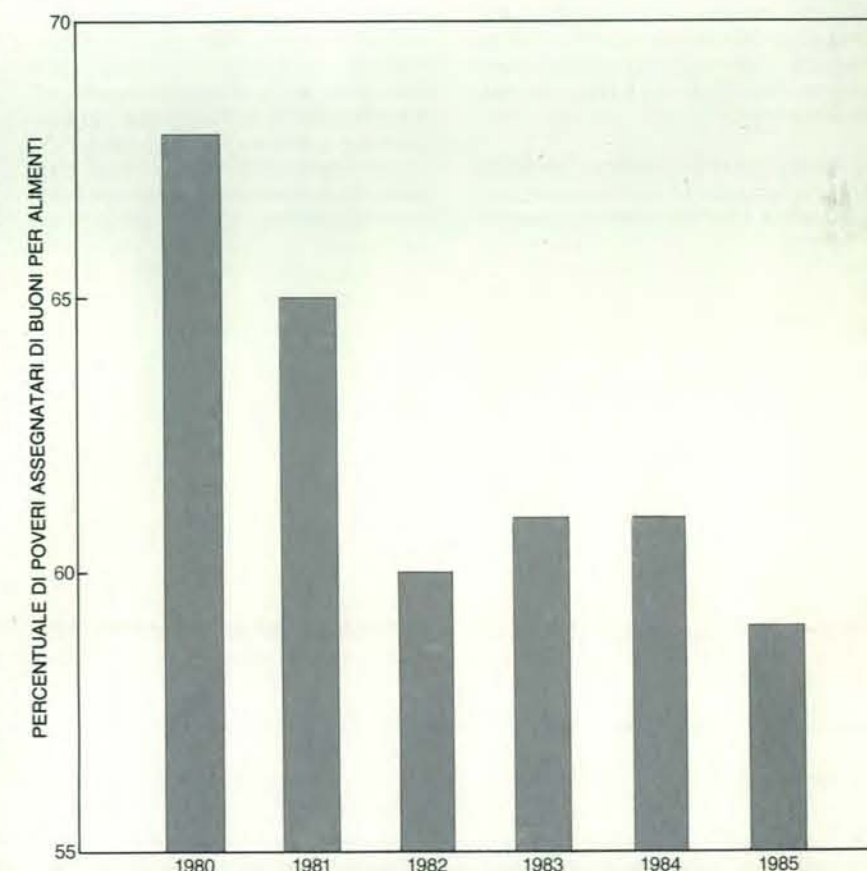
In effetti, recenti ricerche indicano che deficienze funzionali possono essere il risultato di una scarsa nutrizione nell'infanzia anche in assenza di un danno fisico manifesto. Questi studi fanno pensare che l'organismo del bambino si adatti a un'alimentazione inadeguata riducendo il metabolismo prima che compaiano cambiamenti facilmente misurabili nella crescita e prima che si renda manifesta qualsiasi malattia attinente alla nutrizione. Ciò significa che il bambino può non presentare deficienze manifeste, ma essere tuttavia privato di esperienze sociali e cognitive che normalmente favoriscono lo sviluppo.

Gli effetti di una nutrizione insufficiente si possono manifestare anche sugli adulti giovani e di mezza età sia come ridotta produttività sul lavoro sia come inadeguatezza alle richieste imposte dal ruolo sociale. Gli adulti di ogni età sono vulnerabili a infezioni e a varie malattie da carenza associate alla malnutrizione.

Nella vecchiaia i rischi dovuti alla cattiva nutrizione si acuiscono ulteriormente. In età avanzata l'effetto dell'alimentazione sulla conservazione di un buono stato di salute e sulla prevenzione delle malattie è cruciale. Sovente le persone di una certa età hanno bisogno di una dieta particolare a causa della presenza di disturbi diffusi, come ipertensione e diabete. Le malattie da carenza fanno aumentare la necessità di certe sostanze nutritive; ne è un esempio l'accresciuta assunzione di calcio per prevenire l'osteoporosi. Alcune condizioni tipiche dell'età avanzata possono provocare insufficienze nella digestione o nell'assimilazione delle sostanze nutritive, rendendo critica la scelta di cibi di elevato potere nutritivo. Al tempo stesso la vecchiaia pone spesso numerosi ostacoli a una dieta adeguata. Difficoltà nello svolgimento di certe attività, quali l'andare a fare la spesa e il preparare da mangiare, problemi di masticazione e mancanza di appetito dovuta all'isolamento sociale, con conseguente stato depressivo, possono tutti interferire, nella tarda età, con una buona nutrizione.

In questa elencazione degli effetti della fame la situazione dei bambini in tenerissima età merita un'ulteriore considerazione. La mortalità al di sotto del primo anno di vita (cioè il numero di bambini che muoiono nel primo anno di vita per mille nati vivi) è ampiamente accettata come indice significativo dello stato di salute generale di una nazione. Se si usa questo metro, gli Stati Uniti non brillano certamente rispetto ad altre nazioni industrializzate; una mortalità dell'11,2 per mille li pone, infatti, al 18° posto nel mondo. Giappone, Svezia e Regno Unito sono fra i paesi in posizione migliore, ma perfino Spagna e Hong Kong precedono gli Stati Uniti.

Questo dato cela, in realtà, le discrepanze esistenti fra le varie sottopopolazioni degli Stati Uniti. I neonati non bianchi hanno il doppio di probabilità di quelli bianchi di morire nel primo anno di vita. Come ha dichiarato il Children's Defense Fund nel suo rapporto del 1984, *American Children in Poverty*: «Nella capitale la mortalità dei neonati non bianchi supera quella di Cuba e della Giamaica, paesi entrambi molto più poveri. I dati, a livello di Stato, celano problemi di mortalità nel primo anno di vita ancora più gravi nei sobborghi di grandi città. Per esempio, due opuscoli, che riportano i dati di un censimento effettuato a Baltimora, indicano tassi elevatissimi, dell'ordine del 59,5 per mille nati vivi. Questo valore supera la mortalità di Costarica, Panama, Guyana, Trinidad e Tobago relativi al 1981 ed è più del doppio di quella dell'Unione Sovietica.»

Una ulteriore causa di preoccupazione è il fatto che all'inizio degli anni ottanta la mortalità nel primo anno di vita è aumentata, di fatto, in 11 Stati e in molte aree urbane. Inoltre, è possibile che il declino della mortalità nel primo anno di vita negli Stati Uniti nel loro complesso non indichi necessariamente che vengono messi al mondo bambini più sani. Gli Stati Uniti dispongono forse della tecnologia più raffinata al mondo per salvare i bambini nati prematura-

Dal 1980 è diminuito il numero di coloro che usufruiscono dei buoni per alimenti, indicato come percentuale del numero di coloro che vivono a livello di povertà, ossia con un reddito di circa 11 000 dollari all'anno per una famiglia di quattro persone. L'istogramma mette a confronto il numero di coloro che vivono in povertà (*in colore*) e di coloro che ricevono i buoni (*in grigio*).

Il numero di quanti ricevono buoni mensili per alimenti è espresso come percentuale del numero di coloro che sono a livello di povertà o al di sotto. Dal 1980 al 1985, il Governo statunitense ha ridotto di 7 miliardi di dollari gli stanziamenti ed è stato più rigido nelle assegnazioni.

mente o con un peso troppo basso per sopravvivere con le loro sole forze. Questa tecnologia salva una percentuale di bambini nati sottopeso abbastanza elevata, tale da mettere in ombra le reali conseguenze di una nutrizione carente e di cure prenatali inadeguate.

La mortalità nel primo anno di vita non è l'unico evento successivo alla nascita che riveli il mediocre stato di salute e l'insufficiente nutrizione della gestante. La documentazione relativa al peso ridotto alla nascita indica che, al pari della mortalità nel primo anno di vita, esso è distribuito in maniera non uniforme. Il Child Health Outcomes Project, che è stato realizzato presso l'Università del North Carolina, ha rilevato che i bambini negri hanno il doppio delle probabilità dei bambini bianchi di nascere sottopeso. Nel 1982, nella città di New York, dove si dispone di dati sul peso alla nascita a livello di comunità, il centro negro di Harlem aveva la più alta percentuale di bambini nati sottopeso (16,3 per cento), mentre il ricco East Side superiore, abitato in prevalenza da bianchi, aveva la percentuale più bassa (5,4 per cento).

Il ruolo della povertà come causa di un basso peso alla nascita è evidente nei dati che si riferiscono a pazienti poveri, negri e bianchi, ricoverati al Metropolitan Nashville General Hospital. A questo proposito, i due gruppi erano molto vicini (14,4 e 12,9 per cento nel 1981) e i valori fanno pensare che le differenze che si riscontrano nell'esito della gravidanza siano dovute non tanto a qualche fattore biologico legato alla razza quanto piuttosto al reddito e a fattori correlati con la povertà.

Analoghe variazioni socioeconomiche si riflettono nei dati sulla crescita dei bambini. Dei vari indici che caratterizzano lo stato di salute e di nutrizione dei bambini statunitensi, i dati antropometrici (statura, peso e altre dimensioni corporee) sono quelli più ampiamente disponibili. La statura e il peso vengono misurati per consuetudine nel corso dei controlli medici, sicché è possibile avere a disposizione per lo studio una grande quantità di dati.

Questo tipo di ricerca è possibile grazie all'esistenza di un protocollo che consente di confrontare i dati relativi a un singolo bambino. Quando controlla la statura e il peso di un bambino, il medico riporta i valori su un diagramma che mette in relazione la crescita in funzione del tempo. Il diagramma permette anche il confronto con altri bambini della stessa età e dello stesso sesso. Per esempio, una bambina di tre anni classificata nel 40° percentile del peso ha un peso superiore a quello del 40 per cento di tutte le altre bambine della stessa età. Approfondite misurazioni sono state effettuate su bambini di varie culture. Un medico che osserva il diagramma di una bambina di tre anni vuole sapere due cose: primo, se la bambina rientra nell'intervallo atteso per età e sesso; secondo, se la sua crescita segue nel tempo una curva abbastanza regolare.

In genere, i bambini classificati al di sopra del 95° percentile del peso in rapporto alla statura o al di sotto del quinto percentile per quel che riguarda soltanto o il peso o la statura sono tenuti sotto stretto controllo. Può darsi che per alcuni l'andamento sia determinato geneticamente e sia del tutto normale; per altri non è da escludere che la deviazione dalla norma rifletta problemi di natura nutrizionale o altri problemi di salute.

Molti studi condotti negli Stati Uniti su bambini appartenenti a gruppi etnici e razziali diversi hanno rivelato una stretta associazione tra povertà e deficienze di crescita. Per esempio, la Massachusetts Nutrition Survey, un'indagine promossa dall'Assemblea dello Stato del Massachusetts per valutare lo stato di salute dei bambini di famiglie a basso reddito, ha scoperto che il 18,1 per cento dei bambini presi in considerazione mostrava qualche indice di carenza nutrizionale cronica. Il fenomeno era inversamente proporzionale al reddito.

Gli studi che ho citato indicano che la fame e le carenze nutrizionali sono direttamente associate alla povertà e che molte forme di cattiva salute associate a una nutrizione inadeguata sono strettamente collegate al reddito familiare. È probabile che i poveri - proprio perché sono poveri - abbiano un reddito troppo basso per mantenere una dieta adeguata sul piano nutrizionale.

La soglia di povertà, così come viene definita dal Governo degli Stati Uniti, è pari attualmente a poco meno di 11 000 dollari all'anno per una famiglia di quattro persone. Numerosi studi clinici hanno rivelato che la malnutrizione è molto più elevata tra le famiglie che si trovano a quel livello o al di sotto di esso che non tra quelle che sono in condizioni finanziarie migliori. La National Household Food Consumption Survey, pubblicata nel 1982 dall'Agriculture Department, rilevava che più dell'80 per cento di tutti i nuclei familiari la cui spesa per l'alimentazione era analoga a quella del destinatario medio dei buoni per alimenti non riusciva ad avere una dieta con la razione di sostanze nutritive raccomandata dalla National Academy of Science. Il Massachusetts Department of Public Health ha trovato che i bambini che appartengono a famiglie con basso reddito hanno il doppio delle probabilità di essere malnutriti rispetto ai bambini di famiglie con reddito medio. Altri dati sulla salute dei poveri rivelano tassi elevati di malattie correlate con la malnutrizione.

Il reddito, pur non essendo l'unico fattore associato alla cattiva salute, sembra essere quello principale. Qual è precisamente la connessione? Un'ipotesi spesso avanzata è che non sia tanto il reddito di per sé a portare a una maggiore malnutrizione e a provocare malattie tra i poveri quanto piuttosto la mancanza di conoscenze sulla nutrizione, sull'acquisto dei cibi e sul modo di cucinarli.

L'insufficiente informazione in materia di nutrizione è assai diffusa, ma non esistono molte prove a sostegno della tesi secondo la quale si tratta di un fenomeno che riguarda innanzitutto i poveri. Anzi, alcune testimonianze contraddicono quest'ipotesi. Le indagini a livello nazionale sul consumo di cibo hanno rilevato costantemente che i poveri acquistano, in realtà, alimenti più nutrienti del resto della popolazione. A quanto pare i poveri pensano più degli altri a fare scelte giudiziose in quanto debbono sfruttare al massimo i soldi che hanno a disposizione per acquistare cibo. Tanto queste testimonianze quanto il senso comune dimostrano che il recente aumento della fame negli Stati Uniti non può essere attribuito all'ignoranza dei poveri. Coloro i quali avanzano quest'ipotesi dovrebbero spiegare per quale ragione i poveri siano diventati ignoranti così all'improvviso.

La distribuzione di alimenti da parte della Second Harvest National Food Bank Network salì da poco più di un milione di chilogrammi nel 1979 a 45 milioni nel 1985, riflettendo gli sforzi delle organizzazioni private contro la fame. La Second Harvest serve le banche alimentari degli Stati.

La chiara associazione tra povertà e fame fa pensare che, per comprendere il fenomeno della crescita della fame negli Stati Uniti, sia necessario esaminare lo stato delle famiglie a basso reddito. In questo senso la fame è come una qualsiasi malattia di natura infettiva: ha una causa. Sembra che tre siano i fattori responsabili della «malattia» fame. Uno è l'inadeguatezza del sistema di sicurezza sociale della nazione, cioè della rete di assistenza per le persone che vivono in povertà o a un passo dalla povertà. Un altro è il maggior rischio economico dovuto alla recessione. Il terzo è la recente politica governativa, con particolare riferimento ai tagli imposti ai programmi federali riguardanti la nutrizione.

Il sistema di sicurezza sociale americano è debole rispetto a quello di altre nazioni industrializzate. Gli Stati Uniti, per esempio, sono una delle due nazioni industrializzate (l'altra è il Sudafrica) che non abbiano alcun tipo di assicurazione nazionale contro le malattie. Oltre la metà degli Stati nega alle famiglie povere con figli l'assistenza sociale se ci sono entrambi i genitori. Il programma Federal Aid to Families with Dependent Children (AFDC) entra in gioco solamente se la famiglia si scioglie. Molti di quegli stessi Stati non forniscono alcuna copertura opzionale sotto l'egida del Medicaid per la speciale categoria di famiglie che vengono definite «bisognose dal punto di vista medico».

Quasi metà degli americani in povertà non riceve buoni per alimenti. Quelli che li ricevono beneficiano in media di 49 centesimi di dollaro per pasto. A complicare le cose l'Agriculture Department lega l'assistenza con buoni per alimenti a un piano alimentare creato per aiutare i cittadini che si trovano in difficoltà finanziarie temporanee. Quando il piano fu elaborato, i dietisti di quel dipartimento avvertirono che le famiglie che sarebbero vissute a quel livello di spesa per l'alimentazione non avrebbero fruito della razione dietetica giornaliera raccomandata. Per strano che possa sembrare, oggi l'Agriculture Department adotta quel piano come norma cosicché il livello di sussidi raggiungibile con i buoni per alimenti è molto vicino a quello al quale è probabile la malnutrizione.

Nel decennio scorso il reddito medio netto delle famiglie che rientravano nel programma AFDC è sceso quasi del 40 per cento. In altre parole, molto più di un terzo del loro potere di acquisto se n'era andato perché il livello dei sussidi non teneva il passo con l'inflazione. Nessun altro settore dell'economia è rimasto mai tanto indietro, in anni recenti, rispetto all'inflazione.

La disoccupazione negli Stati Uniti è salita dal 6,2 per cento nei primi mesi del 1980 al 10,8 per cento nel 1983 e, anche se da allora è scesa a circa il 7 per cento, i disoccupati sono oggi di più di quelli del 1980. Le cifre, inoltre, non rendono tutta la gravità della situazione.

Oggi riceve il sussidio di disoccupazione una percentuale di gente senza lavoro più piccola che in qualsiasi altro momento della storia recente. Nella recessione del 1975, per esempio, quasi l'80 per cento dei lavoratori disoccupati riceveva un sussidio; la cifra è scesa oggi al 29 per cento, cioè a un livello basso in assoluto per il programma. Le ragioni vanno ricercate nel prolungato ristagno dell'economia e nei cambiamenti intervenuti nella politica federale che rende più rigorosi i requisiti per ottenere l'indennità di disoccupazione. Oggi, in quasi metà degli Stati, se un lavoratore disoccupato, con moglie e figli, ha usufruito di tutti i sussidi, la famiglia non può più ottenerne altri, e neppure l'AFDC o il Medicaid.

Insieme, questi e altri fattori hanno spinto i problemi di natura economica a un livello che non ha precedenti nella storia recente. Un numero sempre maggiore di cittadini si è trovato nelle condizioni di avere bisogno dei programmi federali per l'assistenza contro la fame, ma la risposta del Governo è stata di ridurre l'assistenza a disposizione dei poveri. I tagli imposti ai programmi alimentari fe-

La mortalità nel primo anno di vita (numero di decessi per 1000 nati vivi) indica lo stato di salute di una nazione. Gli Stati Uniti sono al 18° posto fra le 25 nazioni più industrializzate con una popolazione di due milioni e mezzo di abitanti o più. Spesso i piccoli pesano alla nascita meno di 2500 grammi come conseguenza della scarsa nutrizione della madre durante la gestazione.

La povertà e la crescita insufficiente sono correlate. L'istogramma riporta le percentuali di bambini che, nella Massachusetts Nutrition Survey del 1983, furono trovati di bassa statura per l'età. I valori a sinistra indicano il reddito della famiglia come percentuale del livello di povertà.

derali sono stati i più profondi della storia degli Stati Uniti. Due dei programmi alimentari di base (buoni per alimenti e refezione scolastica) hanno subito tagli sostanziali. Altri due (quello inteso a dare da mangiare agli anziani e il programma WIC) sarebbero stati ridotti in maniera pressoché analoga se non fosse intervenuto il Congresso.

Fra i cambiamenti imposti all'assistenza con buoni per alimenti, oltre ai tagli per un totale di 7 miliardi di dollari negli esercizi finanziari 1982-1985, è stato eliminato un adeguamento degli stanziamenti che tenesse conto dell'inflazione nel 1982. Inoltre, le famiglie con un reddito appena al di sopra del livello ufficiale di povertà sono state cancellate dal gruppo degli aventi diritto a meno che non ne facessero parte elementi anziani o invalidi. Parecchi altri cambiamenti hanno ridotto i sussidi fondamentali proprio in un momento in cui l'inflazione, la disoccupazione e la povertà andavano aumentando. Di conseguenza, oggi circa 19 milioni di americani ricevono buoni per alimenti quando il Governo stesso riconosce che le persone che vivono in povertà sono quasi 33 milioni e mezzo.

Nel 1982, circa 23 milioni di bambini hanno usufruito del programma di assistenza mediante refezione scolastica e quello è stato il massimo tentativo per migliorare l'alimentazione dei bambini poveri. In seguito, nello stesso anno, il programma fu ridotto in maniera notevole: furono cambiati i criteri, basati sul reddito, con i quali venivano determinati gli aventi diritto e fu modificata la procedura necessaria per chiedere l'assistenza in modo da rendere più difficile alla gente qualificarsi. In totale, negli esercizi finanziari 1982-1985 furono tagliati cinque miliardi di dollari dai programmi di alimentazione per l'infanzia. L'effetto immediato fu l'esclusione di un milione di bambini dal programma di refezione scolastica. Il che significa che, dati i cambiamenti intervenuti, circa 2700 scuole smisero di offrire il pasto di mezzogiorno, mentre più di 400 000 bambini e 800 scuole furono cancellati dal programma che prevedeva la distribuzione a scuola della prima colazione.

Uno degli importanti sussidi dei programmi scolastici è sempre stato quello di assicurare uno e spesso due pasti nutrienti al giorno, durante il periodo scolastico, ai bambini appartenenti a famiglie a basso reddito. Un gran numero di funzionari scolastici afferma inoltre che quei programmi portano anche benefici di natura sociale ed educativa.

La conversione in legge dei tagli di bilancio per un ammontare di 12 miliardi di dollari nei programmi federali è giunta in un momento in cui la povertà stava toccando il massimo livello dell'ultimo ventennio. Quei tagli hanno tolto ogni sostegno economico a molte persone, quelle povere da sempre e i nuovi poveri, senza che vi sia alcun sistema di sicurezza sociale pronto ad accoglierle.

Il microtubulo come motore intracellulare

Una recente tecnica di microscopia ottica ha dimostrato che questa componente ormai ben conosciuta delle strutture responsabili della forma e della divisione cellulare interviene anche nel trasporto bidirezionale

di Robert Day Allen

I microtubuli sono entrati in epoca relativamente recente nel novero di quegli elementi che sono riconoscibili in tutte le cellule. Trent'anni fa si poteva solo immaginare la presenza di queste fibre sottili. Venticinque anni fa i biologi scoprirono per la prima volta le loro delicate caratteristiche. Quando poi impararono a trovarli, i microtubuli apparvero loro in una sorprendente varietà di funzioni: dal fuso che organizza la divisione cellulare, ai filamenti che avvolgono le cellule del sangue che non si dividono mai. I microtubuli si guadagnarono presto la fama di valide unità strutturali del citoscheletro.

Adesso sembra che la versatilità di questi organelli vada ben oltre le pure e semplici prestazioni strutturali. Mediante l'aiuto di nuove e potenti tecniche computerizzate applicabili al microscopio ottico, siamo riusciti a osservare microtubuli che spostavano con sorprendente velocità vescicole e organelli attraverso il citoplasma. Ciascun sottile filamento, lungo parecchi micrometri e costituito da una proteina chiamata tubulina, può far scorrere gli organelli contemporaneamente in versi opposti; inoltre un microtubulo può muoversi su un vetrino da microscopio anche se è rotto e separato dal suo ambiente cellulare.

I miei collaboratori e io abbiamo tentato di chiarire il meccanismo con il quale i microtubuli si spostano e muovono il loro carico intracellulare. La nostra ricerca e quella di molti altri ha collegato, per esempio, i microtubuli al trasporto che coordina gli scambi tra il corpo della cellula nervosa e le sue terminazioni. Nell'uomo parecchie malattie neurologiche sono da ricollegare ad alterazioni dei microtubuli. Al di fuori del sistema nervoso, fenomeni tanto diversi come i cambiamenti di colore nei pesci e il movimento dei virus all'interno della cellula dipendono dall'attività di trasporto dei microtubuli. La nostra ricerca può così dare risultati utili in campi più vasti di quello dei semplici protozoi acquatici, che aveva destato i miei primi interessi.

Il trasporto di particelle, che chiamiamo trasporto microtubulo-dipendente, fu studiato parecchio tempo prima della scoperta degli stessi microtubuli. La più chiara descrizione storica di questo trasporto è quella di Joseph Leidy, un naturalista e microscopista vissuto nel XIX secolo. Leidy trovò in uno stagno vicino a casa un curioso animaletto dotato di una mirabile rete di particolari pseudopodi («falsi piedi») detti filopodi. Questo piccolo organismo, che egli chiamò *Gromia*, era un foraminifero d'acqua dolce planctonico, appena visibile a occhio nudo. All'interno dei filopodi, Leidy osservò un flusso dinamico che descrisse nel 1879 nella sua relazione esposta alla US Geological Survey: «Le estensioni pseudopodali di *Gromia* sono costituite da un protoplasma granuloso chiaro, con granuli più grossi che si muovono incessantemente lungo i filamenti in versi opposti tranne che in quelli più sottili.»

Successivamente, altri scienziati fecero eco alla particolareggiata descrizione che Leidy aveva fatto dei filopodi di *Gromia*, ma aggiunsero ben poco a quanto era già stato spiegato degli effetti di quell'insolito flusso. Per esperienza, sappiamo che il flusso all'interno di una struttura chiusa può avvenire in un solo verso, secondo un gradiente di pressione. Ciò che nelle reti di filopodi dei foraminiferi ha affascinato me e altri colleghi era il fatto che gli organelli potessero muoversi all'interno di un singolo filopodo a velocità differenti e in entrambi i versi. In qualche modo i foraminiferi contravvengono alle regole.

Per svelare i segreti di *Gromia* ci volle un poco di tempo. Nel 1957, mi recai da Zach M. Arnold all'Università della California a Berkeley per studiare il movimento dei filopodi in *Allogromia laticollaris*, un foraminifero marino che Arnold aveva fatto crescere in coltura e al quale aveva attribuito quel nome. Questo organismo e un altro di una specie affine, *Allogromia* sp. N. D., sarebbero diventati il materiale di partenza per importanti esperimenti che avrebbero dovuto collegare il movimento delle particelle ai microtubuli. Essi sono simili a *Gromia*, ma a differenza del foraminifero di Leidy, sono sufficientemente numerosi da permettere una ricerca e possono così essere utilizzati come validi modelli sperimentali.

Frequentai il laboratorio di Berkeley sei anni prima che i microtubuli fossero scoperti, quando le tecniche di microscopia erano ancora troppo primitive per poter osservare più particolareggiatamente quello che Leidy aveva già descritto. La ricerca era lenta e monotona. Tuttavia, potemmo considerare alcune semplici ipotesi, prevedendo i tipi di movimento e verificando le previsioni con semplici osservazioni. Nel 1959, affrontando il problema in questi termini, Theodore L. Jahn e Robert A. Rinaldi dell'Università della California a Los Angeles suggerirono che le particelle in movimento fossero catturate da fibre di gel «simili a millepiedi» capaci di scorrere l'una sull'altra spostandosi a vicenda tramite «minuscoli arti».

L'esperienza che ho acquisito lavorando su *Allogromia* e alcuni foraminiferi affini presso il Laboratoire de Biologie marine di Roscoff, in Francia, mi ha convinto che il modello di Jahn e Rinaldi era troppo semplice. Nei filopodi più sottili di quei protozoi le particelle si muovono più velocemente verso la punta che non nel verso opposto. Ho proposto allora un modello nel quale i filopodi più piccoli contenevano fibre di gel piegate in due, a forma di V, che si contraevano nella zona di piegatura vicino alla punta. Questo modello riusciva a spiegare perché il movimento fosse più rapido in un verso che non nell'altro, ma ricerche successive hanno dimostrato che io non ero più vicino alla verità di quanto lo fossero i ricercatori dell'Università della California a Los Angeles.

Nel corso delle mie prime indagini, però, avevo notato che la rete dei filopodi appariva molto luminosa quando era osservata al microscopio in luce polarizzata. Il fenomeno suggeriva la presenza di fibre submicroscopiche orientate. Qualunque fosse la natura di queste fibre, esse dovevano essere molto delicate o comunque instabili: era sufficiente il minimo cambiamento nelle condizioni esterne per far sì che si trasformassero in goccioline di citoplasma contenute in involucri membranosi.

È stata proprio questa fragilità a rendere vani per quasi dieci anni i miei tentativi di dimostrare che i microtubuli erano numerosi nelle reti di filopodi dei foraminiferi. Nel 1963, David B. Slautterback della Medical School dell'Università del Wisconsin ha identificato per primo i microtubuli nell'idra, un piccolo celenterato di acqua dolce. Nello stesso anno Myron C. Ledbetter e Keith R. Porter dell'Harvard University hanno notato alcuni microtubuli nelle cellule vege-

Nel neurone i microtubuli (*in giallo*) coordinano il trasporto di vescicole (*in blu* e *cremisi*) e organelli (*in viola*) attraverso il «corridoio» che collega il corpo cellulare alle terminazioni.

Gromia, un protozoo d'acqua dolce affine alle amebe, fornì la prova del movimento bidirezionale oltre un secolo fa. Il microscopista Joseph Leidy, a cui si deve questo disegno del 1879, notò un flusso di «granuli», che si spostavano in due versi all'interno dei lunghi e sottili filopodi che si allungano come tentacoli dal corpo dell'animale. Il flusso bidirezionale all'interno di una singola struttura chiusa era un evento inspiegabile e diede luogo a varie ipotesi che indussero l'autore ad approfondire le ricerche. A occhio nudo _Gromia_ non è più grande di un granello di sabbia.

tali. Abbiamo avuto subito il sospetto che i microtubuli costituissero gli elementi di sostegno di *Allogromia*. Ma solo nel 1971 Samuel M. McGee-Russell e io, lavorando alla State University di New York ad Albany, abbiamo trovato il sistema per migliorare quei metodi che, nell'allestire i preparati, distruggevano i microtubuli prima che potessero essere visti al microscopio elettronico.

Allo stesso modo trascorsero altri dieci anni prima che riuscissimo a documentare la partecipazione diretta dei microtubuli nel movimento di particelle. Nel 1981, due successi ottenuti contemporaneamente nel mio laboratorio, al Dartmouth College, hanno permesso di osservare con la massima nitidezza quelle fragili fibre. Jeffrey L. Travis, uno studente che svolgeva ricerche nel mio laboratorio sulla mobilità di *Allogromia*, aveva messo a punto un metodo che permetteva la crescita di questo organismo su vetrini coprioggetti. Travis trattava vetrini con sostanze dotate di carica positiva, come la poli-L-lisina, così da far aderire i filopodi al vetrino e da farli distendere in lamine ampie e sottili, ideali per osservare il movimento degli organelli. Egli aveva anche preparato campioni interi al posto di sezioni e questi hanno rivelato per la prima volta al microscopio elettronico il complesso sistema di circuiti presente nei microtubuli.

Nel frattempo la mia casuale e fortunata scoperta di quella che sarebbe stata chiamata microscopia AVEC (microscopia di Allen con contrasto aumentato mediante telecamera) ci permise di vedere strutture troppo piccole per poter essere osservate al microscopio ottico tradizionale. Abbiamo così potuto studiare i singoli microtubuli anche se il loro diametro era di soli 25 nanometri (25 milionesimi di millimetro). Grazie a questo tipo di microscopia, Travis e io abbiamo osservato per la prima volta l'interazione tra particelle e tubuli. Abbiamo anche notato che era sufficiente un breve contatto fisico con i microtubuli per determinare lo spostamento di particelle ferme. Travis aveva poi constatato che i microtubuli si spostavano con movimenti assiali e laterali all'interno della rete dei filopodi. Non siamo invece riusciti a stabilire se si muovevano attivamente o se venivano spinti passivamente, né se le particelle si spostavano lungo un singolo microtubulo.

Abbiamo pensato che la causa di queste limitazioni fosse il nostro modello sperimentale e così abbiamo puntato l'attenzione su sistemi cellulari diversi da *Allogromia* analizzabili con la microscopia AVEC. Ci siamo così occupati del trasporto di organelli lungo l'assone, cioè il prolungamento principale della cellula nervosa, in cui si propaga l'impulso. Nei neuroni di pollo si sapeva che il trasporto assonico avveniva nei due sensi con una velocità di circa 200 millimetri al giorno, una velocità discreta in scala microscopica. Questo trasporto rapido è stato descritto per la prima volta nel 1965 da un mio allievo, William Burdwood. Quando egli riferì questa scoperta, la sua risonanza fu a dir poco sensazionale: nei quindici anni successivi, per spiegarla, presero corpo svariate ipotesi, ma poiché le prove sperimentali erano scarse, nessuna teoria prese il sopravvento.

Nell'estate del 1981, mentre lavoravo presso il Marine Biological Laboratory di Woods Hole, decisi di verificare se la microscopia AVEC poteva dare qualche suggerimento per il trasporto nell'assone gigante di calamaro, un materiale ideale per la sperimentazione. L'assone faceva parte di un neurone grande e trasparente che si prestava bene ai miei fini, a parte il fatto che nessuno aveva notato in esso alcuna traccia di movimento. Per le mie ricerche chiesi anche la collaborazione di parecchi colleghi che avevano maturato una buona esperienza nello studio dei neuroni degli invertebrati.

Senza l'attrezzatura per la microscopia AVEC, l'osservazione al microscopio a contrasto di interferenza differenziale è stata piuttosto deludente. Abbiamo potuto vedere solo confusamente il contorno dell'assolemma (la membrana che avvolge l'assone) e un sottile strato di cellule connettivali che lo ricoprivano, ma non siamo riusciti a osservare alcuna struttura in movimento. Poi siamo ricorsi alle tecniche AVEC e abbiamo utilizzato il sistema video. Immediatamente sono apparse sullo schermo innumerevoli particelle che si muovevano freneticamente in entrambi i versi lungo una sorta di filamenti. La nostra emozione è stata tale che l'abbiamo trasmessa a tutta la comunità di Woods Hole.

In luglio abbiamo potuto eseguire le prime registrazioni del flusso di particelle lungo l'assone gigante di calamaro e, in base alla loro forma, abbiamo potuto riconoscerne parecchie. I mitocondri, grossi e allungati depositari dell'energia chimica contenuta nell'adenosintrifosfato (ATP), si spostavano in entrambi i versi con un movimento a scatti, discontinuo. Le particelle sferiche più piccole che si dirigevano verso il corpo della cellula erano corpi multivescicolari che trasportavano frammenti di membrana «in sovrappiù». Le particelle di dimensione minima e appena visibili si spostavano ininterrottamente in massa, come fitti banchi di piccoli pesci, verso la terminazione sinaptica, cioè verso l'estremità dell'assone. Esse erano i precursori delle vescicole sinaptiche contenenti i neurotrasmettitori, i quali vengono rilasciati quando la cellula nervosa viene stimolata, e non erano mai state viste prima di allora con il microscopio ottico.

Quando l'entusiasmo si placò, fummo in grado di valutare i risultati ottenuti. I nostri esperimenti avevano perlomeno stabilito i versi in cui si muovevano le differenti particelle e avevano dimostrato che le particelle più grosse si spostavano più lentamente di quelle piccole. Sembrava che avessimo ottenuto la prima prova dell'esistenza di un movimento continuo di particelle nelle cellule animali. Un tale tipo di trasporto era inaspettato, perché fino ad allora questo movimento era stato osservato solo nelle correnti citoplasmatiche delle cellule vegetali. Lo spostamento di particelle nelle cellule animali era sempre stato caratterizzato da una certa discontinuità.

Che cosa provocava il movimento? Non sembrava che le particelle, contrariamente a quanto osservato nelle cellule vegetali, si muovessero trascinate dalle correnti citoplasmatiche. Alcuni ricercatori avevano proposto che fosse la carica elettrica associata all'impulso nervoso ad alimentare in qualche modo il trasporto lungo l'assone. Ma quando abbiamo stimolato elettricamente l'assone, il processo non ha subito variazioni. Le particelle hanno continuato a muoversi seguendo invariabilmente quelle «strutture lineari» che assomigliavano moltissimo ai microtubuli che avevamo visto in *Allogromia*.

Meditando sulle nostre osservazioni collettive ho iniziato a pensare che i microtubuli potessero partecipare al trasporto di particelle sia nei foraminiferi, sia nei neuroni. Bisognava però considerare il fatto che gli assoni contenevano altre «strutture lineari», come i neurofilamenti e l'actina *F*. All'epoca (1981) la teoria prevalente era che l'elemento maggiormente interessato nel trasporto lungo gli assoni fosse proprio l'actina *F* (si veda l'articolo *Il trasporto di sostanze nelle cellule nervose* di James H. Schwartz in «Le Scienze» n. 142, giugno 1980).

Con una risoluzione di circa 100 nanometri nei casi migliori, non eravamo in grado di dire quali fossero le strutture lineari, quale fosse il loro grado di addensamento e che rapporti esse avessero esattamente con le particelle trasportate. Pensavamo di poter continuare a rispondere a questi interrogativi separando l'assone, isolandone i vari elementi e particelle e osservandone le interazioni tra questi ultimi.

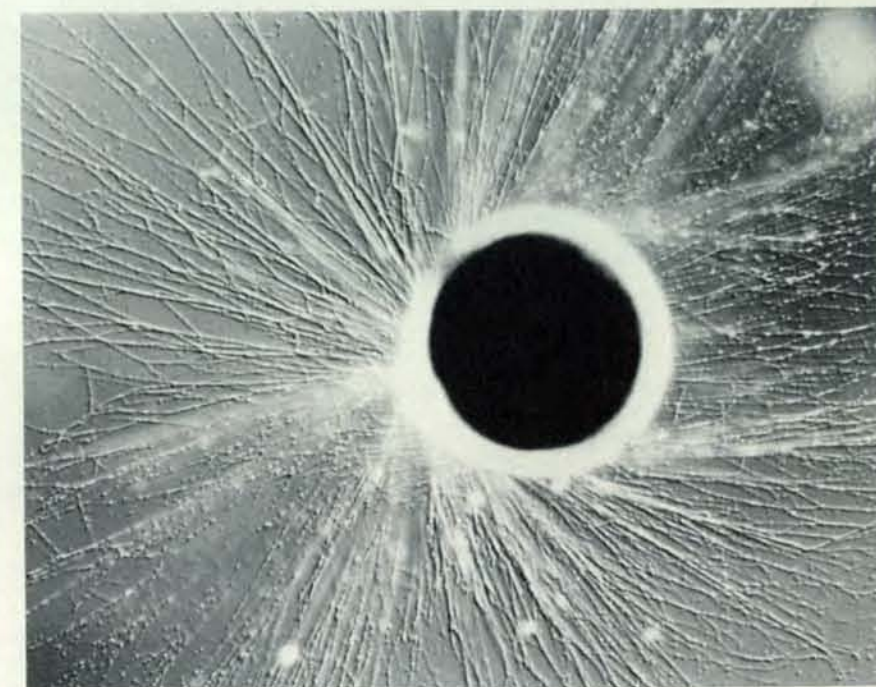
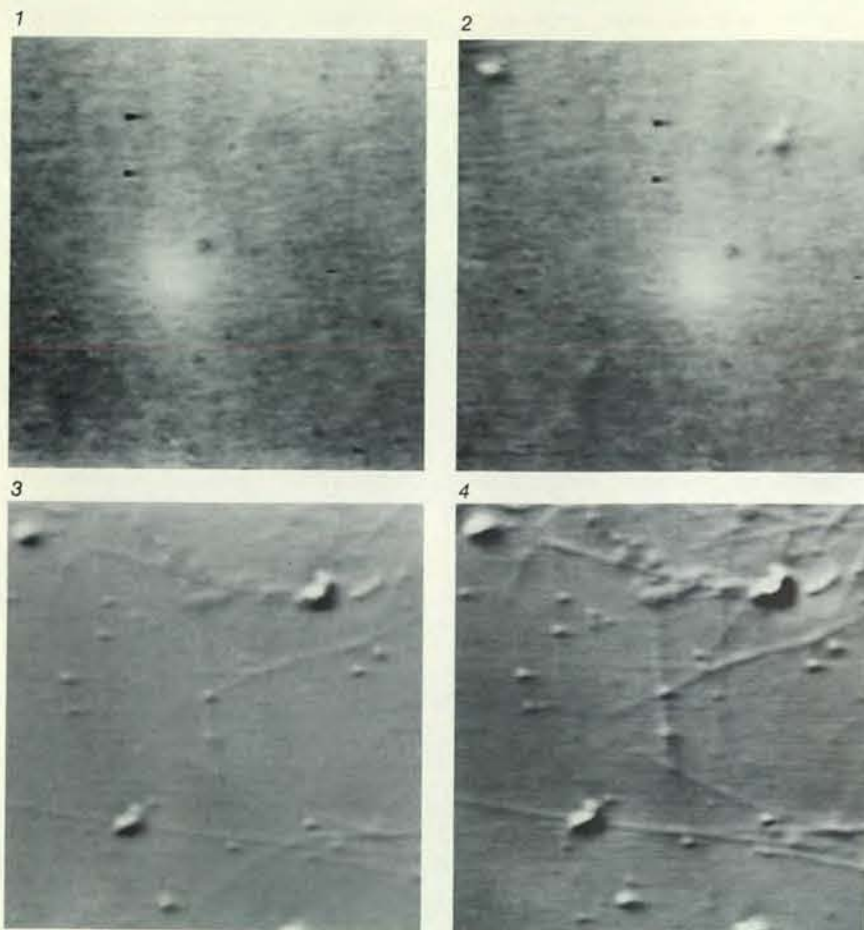
Raymond J. Lasek della School of Medicine della Case Western Reserve University, una personalità nel campo del trasporto assonico, e Scott T. Brady, un ricercatore che si stava specializzando nel suo laboratorio, hanno suggerito di usare una tecnica messa a punto 40 anni prima a Woods Hole da Richard S. Bear, Francis O. Schmitt e John Z. Young. Questi ricercatori avevano trovato il modo di estrarre l'assoplasma, o citoplasma dell'assone, da un singolo assone. L'assoplasma fuoriesce come un cilindro gelatinoso, che non si fonde e conserva in questo modo l'integrità del suo contenuto.

Brady ha adattato questa tecnica al problema che lo interessava e ha ottenuto immediatamente risultati positivi. Il trasporto di particelle è continuato indisturbato nell'assoplasma, privato della membrana cellulare, e, in presenza di una soluzione di sali, di amminoacidi e di ATP (la molecola ad alto contenuto di energia), è rimasto attivo per ore, lasciando ampio spazio alla sperimentazione. Ho suggerito di «rimescolare» l'assoplasma estratto in modo da verificare se il meccanismo di trasporto dipendesse dall'integrità della sua struttura. Il rimescolamento non ha interrotto il movimento degli organelli, ma lo ha semplicemente disorganizzato. Nel frattempo le particelle hanno continuato a seguire la traccia lasciata dalle «strutture lineari». Ciò mi ha convinto sempre più che i microtubuli mantenevano la loro attività e ha avallato la mia convinzione che svolgessero un ruolo nel trasporto delle particelle.

Queste convinzioni hanno poi trovato conferma nei risultati della ricerca condotta nel mio laboratorio di Dartmouth da John H. Hayden. Egli ha separato dall'occhio di anfibio i cheratinociti, cellule connettivali specializzate, con zone periferiche estremamente sottili. All'interno di questi margini siamo riusciti a vedere nelle cellule vive, mediante microscopia AVEC, particelle che si spostavano nei due versi opposti, lungo quelle che sembravano essere «strutture lineari» uguali.

Hayden ha verificato la composizione di queste strutture con anticorpi antitubulina preparati da R. D. Goldman della Medical School della Northwestern University. Questi anticorpi si sono legati alle fibre dimostrando per la prima volta che le tracce filamentose che venivano seguite dalle particelle erano costituite da tubulina e quindi erano microtubuli.

Bisognava ora controllare se un singolo microtubulo poteva muoversi in un solo verso o entrambi. In una seconda serie di esperimenti, che richiedevano la massima capacità tecnica nell'allestire il preparato, Hayden, con l'aiuto della videomicroscopia, è riuscito a registrare innanzitutto gli spostamenti delle particelle in entrambi i versi e quindi a fissare gli stessi cheratinociti per l'esame al microscopio elettronico. Confrontando le due immagini egli ha potuto osservare che singoli microtubuli convogliavano le particelle in entrambi i versi. Questa è stata la prima prova conclusiva che un singolo microtubulo poteva essere il responsabile di quell'inspiegabile movimento a doppio senso, che sembrava sfidare le leggi della natura in *Allogromia*. Pensavamo, infatti, che i microtubuli dovessero essere implicati con tutta probabilità in ciascuno dei sistemi da noi studiati: foraminiferi, cellule tissutali e neuroni.

I progressi compiuti da Hayden hanno posto altri interrogativi. Nelle cellule ogni movimento orientato necessita di energia chimica e di un certo tipo di «enzima generatore di forza», che trasformi l'energia chimica in energia meccanica. Nelle cellule, la più comune fonte di energia chimica è l'adenosintrifosfato (ATP) che libera l'energia immagazzinata quando reagisce con una molecola d'acqua nel processo di idrolisi. Noi abbiamo avanzato l'ipotesi che l'e-

La microfotografia, ottenuta con un normale microscopio a contrasto di interferenza differenziale, mostra la complessa rete di filopodi che prende origine dal corpo cellulare di _Allogromia_. La prima prova diretta del trasporto di particelle attraverso i microtubuli è stata osservata proprio in queste ramificazioni trasparenti. _Allogromia_ ha un diametro di mezzo millimetro.

La microscopia di Allen a contrasto aumentato mediante l'impiego di telecamera (per questo detta AVEC, dall'inglese *Allen video-enhanced contrast*) funziona amplificando le differenze di luminosità e riducendo la luminosità media dell'immagine televisiva finale. Ecco come si applica al microscopio ottico: per prima cosa, si aumenta l'ingrandimento applicando uno zoom al microscopio (in questo caso un microscopio a contrasto di interferenza differenziale). In seguito, si mette a punto il microscopio in modo da far passare più luce; così facendo, si aumenta il rapporto segnale visivo-sfondo e si rende ottimale l'uso della telecamera. Ciò rende, tuttavia, difficile la visione dell'immagine perché l'occhio rimane abbagliato dalla luce diffusa. L'immagine *1* si ottiene quando si amplifica la componente modulata della tensione del segnale televisivo, aumentando le differenze di luminosità, e quando si sottrae una corrente continua, che rappresenta la luminosità di fondo. L'aumento di contrasto rende visibili alcune strutture del preparato (come, per esempio, le due vescicole nella metà superiore della figura), ma evidenzia anche la polvere e le imperfezioni dell'obbiettivo, che rendono scuro il preparato. Poiché questo inconveniente rimane anche sfuocando l'immagine (*2*), lo si può eliminare sottraendolo digitalmente a ogni fotogramma che giunge al video e rendendo così più nitida l'immagine (*3*). Quest'ultima risulta però poco contrastata. Si può aumentare il contrasto digitalmente e si può migliorare il rapporto segnale visivo-sfondo facendo la media del segnale di alcuni fotogrammi (*4*).

nergia necessaria al trasporto microtubulo-dipendente fosse fornita dall'ATP in quanto il movimento delle particelle si arrestava quando i microtubuli ne venivano privati. Ma volevamo saperne di più sugli enzimi generatori di forza: di che cosa fossero fatti, come funzionassero e se fossero associati con i microtubuli, le particelle o tutti e due. Per ricavare le risposte dovevamo isolare ed esaminare i singoli microtubuli.

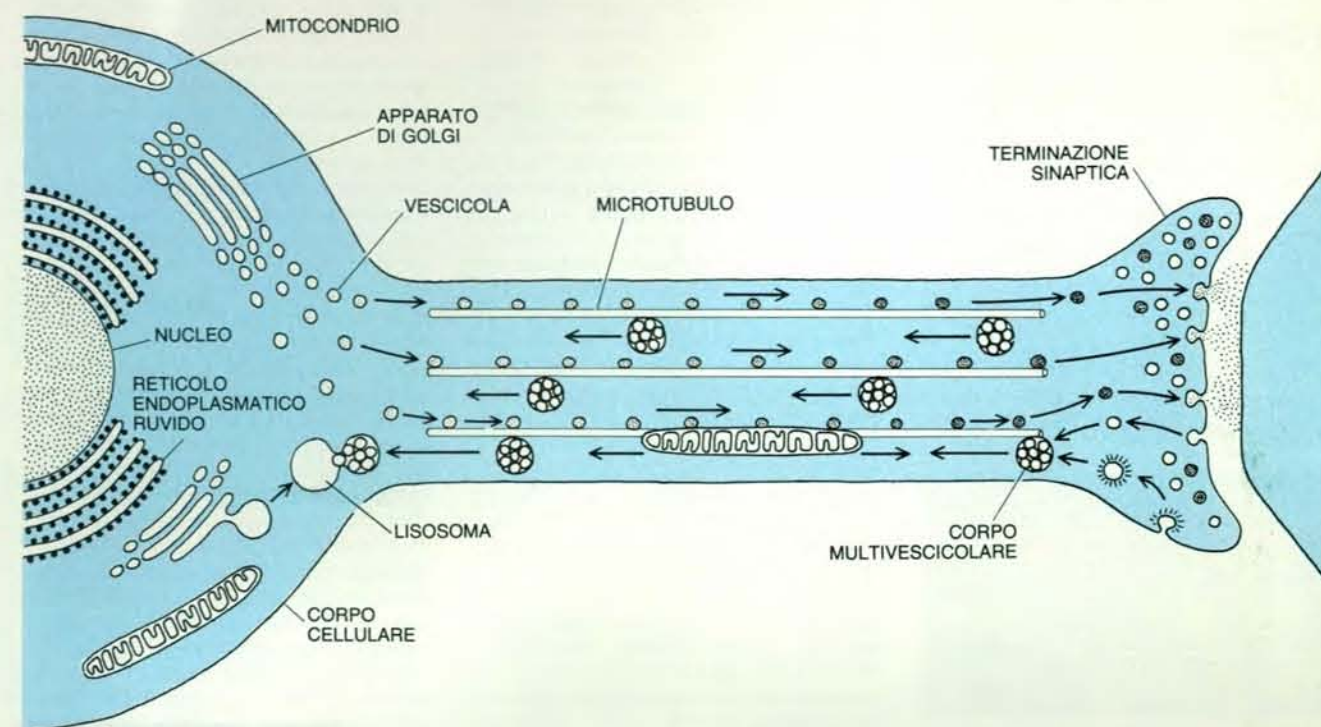
Nell'estate del 1983, ancora a Woods Hole, abbiamo ripreso gli esperimenti sull'assoplasma di calamaro. Siamo riusciti a «diluire» l'assoplasma in modo che i microtubuli si separassero gli uni dagli altri. Osservando poi i microtubuli isolati mediante microscopia AVEC abbiamo potuto notare che le particelle si spostavano come Hayden aveva osservato nelle cellule di tessuti intatti. Abbiamo inoltre constatato che i microtubuli molto lunghi (più di 100 micrometri) potevano cambiare lentamente aspetto durante il trasporto delle particelle assumendo strutture aggrovigliate oppure variamente annodate o a forma di stella.

Sfortunatamente, questi campioni di microtubuli che si agitavano convulsamente erano così lunghi che andavano oltre il campo di osservazione del microscopio e pertanto non potevamo dire se entravano in collisione con oggetti inanimati e si piegavano in seguito all'urto oppure se si deformavano a causa del contatto con le particelle, con la superficie del vetrino o con altri microtubuli. In compagnia di Lasek, Brady e Dean Martz, anch'egli della Case Western Reserve University, avevo osservato che, nell'assoplasma di calamaro isolato dall'assolemma, i mitocondri di forma allungata cambiavano aspetto, il che mi ha suggerito che porzioni della membrana esterna fossero tirate in direzioni diverse in seguito al contatto con due, tre o un numero maggiore di microtubuli vicini. Forse un simile tiro alla fune avveniva anche tra gli stessi microtubuli.

Non abbiamo potuto stabilire quale fosse, tra le varie possibilità, la causa di questo fenomeno finché non siamo riusciti a frantumare i microtubuli in segmenti più corti con una tecnica di cui ci siamo impadroniti solo verso la fine dell'estate. Un anno dopo, in collaborazione con Dieter Weiss, che era venuto a Woods Hole dal Zoologische Institut dell'Università di Monaco di Baviera, ho registrato il movimento di questi frammenti. Abbiamo totalizzato ben 45 ore di riprese su videocassetta, accumulando registrazioni a velocità normale e accelerata 10 volte con la tecnica della ripresa intervallata detta anche videoregistrazione ultraveloce. Le videocassette ci hanno rivelato l'esistenza di un microcosmo inaspettatamente vivace, vario e movimentato.

Ma i reperti più spettacolari li abbiamo trovati quando siamo riusciti a frammentare i microtubuli, che di solito sono lunghi qualche centinaio di micrometri, in segmenti di lunghezza variabile da 25 micrometri (il diametro del campo visivo del nostro microscopio) a valori inferiori al micrometro. Siamo rimasti sorpresi nel vedere questi corti segmenti di microtubuli scivolare liberamente attraverso il vetrino. Ciascuno si muoveva solo in un verso, non retrocedeva mai e avanzava in modo rettilineo a meno che non urtasse contro qualche ostacolo, come un organello rimasto attaccato al vetrino. Poteva rimbalzare o, se urtava qualche ostacolo frontalmente, spostarsi con una serie regolare di movimenti simili a quelli della coda di un pesce.

Questi movimenti a spirale avvenivano sia nel caso che i segmenti trasportassero particelle sia nel caso che fossero liberi. Del tutto inattesi, essi erano estremamente interessanti sotto l'aspetto visivo, ma risultarono egualmente importanti per la comprensione degli enzimi generatori di forza. Qualche anno prima altri ricercatori avevano stabilito come costruire microtubuli partendo da subunità di tubulina purificate (si veda l'articolo *I microtubuli* di Pierre Dustin in «Le Scienze» n. 146, ottobre 1980). Poco dopo le nostre indagini, Thomas S. Reese del National Institute of Neurological and Communicative Disorders and Stroke, assieme ai suoi collaboratori, ha dimostrato, lavorando a Woods Hole, che i microtubuli così costruiti non potevano scorrere se non veniva aggiunta al mezzo una frazione solubile dell'assoplasma di calamaro. Questo dato sperimentale ci ha convinti che il sito d'azione dell'enzima generatore di forza fosse sulla superficie del microtubulo e che, di tale enzima, dovesse esistere una riserva nel citoplasma. I risultati della ricerca hanno anche suggerito che l'enzima è solo debolmente associato al microtubulo e non strettamente accoppiato a esso. Pertanto abbiamo definito l'enzima generatore di forza «proteina che interagisce con il microtubulo».

In che modo, allora, i nostri dati concordavano con le opinioni già consolidate sul meccanismo del trasporto assonico? David S. Smith dell'Università di Miami e altri ricercatori avevano dimostrato, in fotografie al microscopio elettronico, che i microtubuli possedevano «braccia» o ponti trasversali. Era anche stato dimostrato che negli assoni vescicole e organelli erano connessi ai microtubuli mediante ponti trasversali di lunghezza variabile. Nel 1972 Sidney Ochs della School of Medicine dell'Università dell'Indiana aveva supposto che questi ponti trasversali fossero costituiti da una proteina in grado di spostare l'organello in seguito a un cambiamento conformazionale causato da una reazione dipen-

Il trasporto assonico, che avviene lungo i microtubuli, permette un rapido scambio di sostanze tra il corpo e la terminazione sinaptica in corrispondenza della quale la fibra della cellula nervosa si connette con la cellula bersaglio. Le vescicole contenenti i neurotrasmettitori sono sintetizzate dall'apparato di Golgi e trasportate verso la sinapsi. In corrispondenza della terminazione sinaptica, la membrana in sovrappiù è inglobata nei corpi multivescicolari, che ritornano al corpo cellulare dove i lisosomi li degradano. I mitocondri si muovono in entrambi i sensi, soddisfacendo il fabbisogno energetico della cellula. Il trasporto delle vescicole è continuo, quello dei mitocondri intermittente. È chiaro che un singolo microtubulo è in grado di sostenere il trasporto bidirezionale, ma non si conosce ancora il meccanismo che ne è alla base.

I movimenti ondulatori e di scorrimento dei microtubuli isolati suggeriscono la presenza di qualche «enzima generatore di forza» associato alle fibre stesse. Il verso in cui si spostano i microtubuli dipende dalla posizione iniziale, dato che ogni microtubulo può muoversi solo in un verso. Quando il microtubulo incontra un ostacolo, per esempio un organello, invece di continuare a scorrere comincia a raggomitolarsi su se stesso o a contorcersi in spirali. Il moto non è lento: queste immagini sono state ottenute con la tecnica AVEC a intervalli di 18 secondi.

Questa serie di fotogrammi ottenuti con la tecnica AVEC mostra il movimento di particelle in versi opposti sullo stesso microtubulo. Una grossa particella, presumibilmente un corpo multivescicolare, scende lungo la concavità del microtubulo a forma di S e sale dalla parte opposta. Nel terzo fotogramma una particella più piccola e di forma allungata aderisce dove la convessità del microtubulo è massima per poi scivolare lungo il lato sinistro. Questa serie di immagini è stata eseguita con scatti ogni 13 secondi e con un ingrandimento di 9000 volte.

dente dall'energia. Alcuni ricercatori hanno supposto che forse le proteine coinvolte potevano essere l'actina e la miosina, note per la loro capacità di provocare la contrazione muscolare, ma hanno previsto che su ciascun filamento il movimento avvenisse gradualmente e procedesse solamente in un verso, mentre noi avevamo osservato su ciascun microtubulo un movimento continuo e bidirezionale.

I nostri primi tentativi di rendere visibili le braccia laterali hanno incontrato un ben scarso successo, mentre Brady e Lasek conducevano esperimenti che davano adito a qualche speranza. Per chiarire il meccanismo di trasporto dei microtubuli essi hanno usato un analogo dell'ATP (AMP-PNP). Questa molecola sostituisce l'ATP nel sito attivo dell'enzima generatore di forza, ma non può essere idrolizzata e non può quindi fornire l'energia indispensabile per il movimento. Brady e Lasek sapevano che l'AMP-PNP causava l'adesione delle vescicole e degli organelli ai microtubuli e ne hanno dedotto che, probabilmente, esso rendeva stabili anche le braccia laterali. Trattando l'assoplasma che sporgeva dalla membrana con l'analogo dell'ATP Brady è riuscito a fotografare queste braccia al microscopio elettronico. Da preparati analoghi, lo stesso ricercatore, che ora lavora allo Health Science Center dell'Università del Texas a Dallas, ha isolato un'ATPasi, enzima che idrolizza l'ATP e che sembra avere buone probabilità di svolgere il ruolo di enzima generatore di forza. Il gruppo di Reese, lavorando a Woods Hole, ha dato alla ricerca la stessa impostazione di Brady, ma è giunto a una proteina diversa.

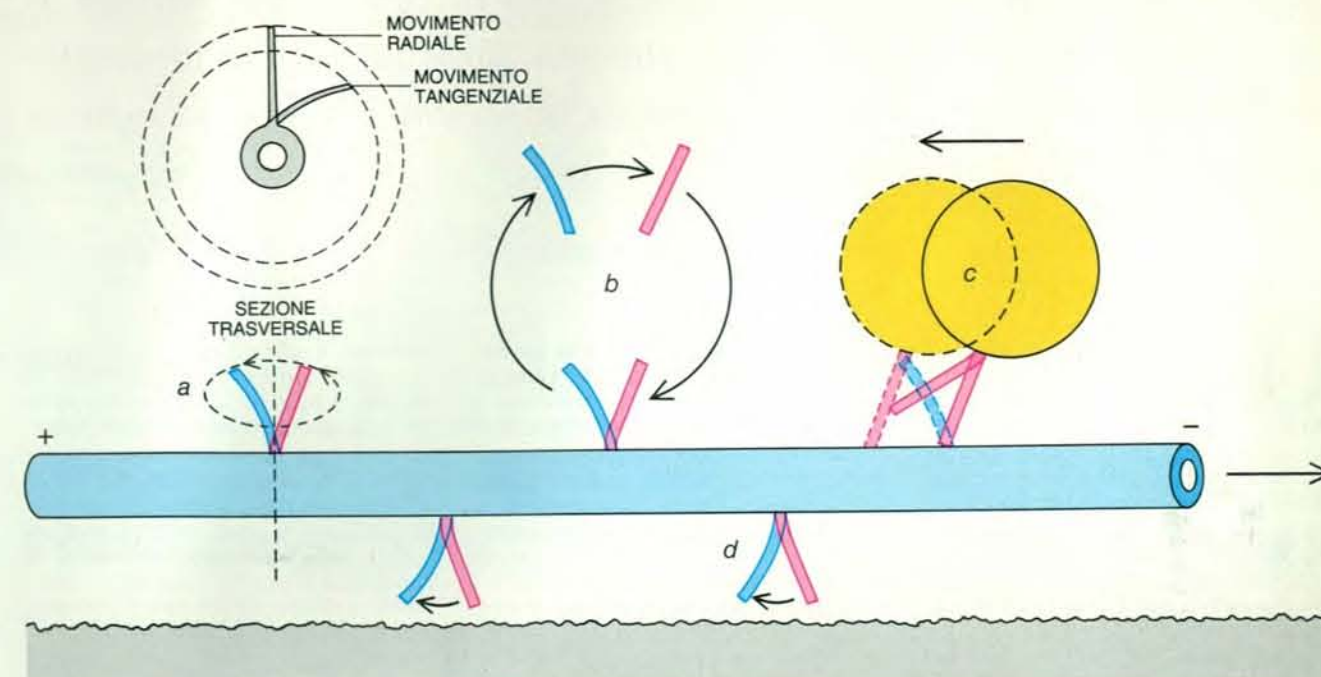
Anche il punto di partenza delle braccia laterali era in discussione. Le braccia laterali, o ponti trasversali, si dipartono realmente dai microtubuli, o dagli organelli e dalle vescicole, oppure sono disperse nel citoplasma? Anche in questo caso ricercatori diversi hanno ottenuto risultati differenti. È chiaro, invece, che in certe circostanze sia i microtubuli, sia il loro carico possono legarsi alle braccia laterali e così abbiamo cominciato a pensare in che modo le braccia laterali e l'enzima generatore di forza possano interagire tra loro per indurre il trasporto microtubulo-dipendente.

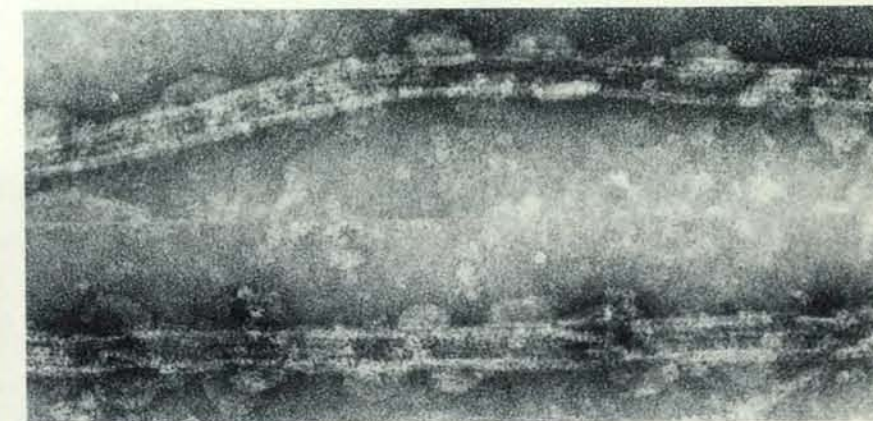
Nel 1985, durante il mio ultimo soggiorno a Woods Hole, in collaborazione con Weiss, George M. Langford della School of Medicine dell'Università della North Carolina, e Dieter Seitz-Tutter, che, con Weiss, lavora oggi al Politecnico di Monaco di Baviera, ho sviluppato un modello della «attività innata del microtubulo», che presuppone che ciascun braccio laterale contenga come enzima generatore di forza l'ATPasi, la quale, idrolizzando l'ATP, induce un cambiamento conformazionale del braccio laterale, che passa così da uno stato «attivo» a uno stato «inattivo» o «di riposo». Mentre cambia conformazione, il braccio laterale o forma ponti trasversali momentanei tra una vescicola e un microtubulo o interagisce con una superficie vicina. La forza, legata al cambiamento conformazionale del braccio laterale, generato dall'idrolisi dell'ATP, viene utilizzata o per spostare una vescicola lungo il microtubulo, o per spostare il microtubulo lungo una superficie vicina.

Sia che le braccia laterali si dipartano dai microtubuli, dalle vescicole e dagli organelli, sia che si muovano liberamente nel citoplasma, noi pensiamo che per innescare il cambiamento conformazionale, sia necessario il contatto con un microtubulo, dato che non sembra che le vescicole si spostino senza avere un contatto con la superficie dei microtubuli. Il modello che abbiamo sviluppato presuppone anche che le braccia laterali subiscano un'attivazione prima di attaccarsi ai microtubuli. Il ciclo di cambiamenti conformazionali può essere analogo al meccanismo actina-miosina che avviene nella contrazione muscolare e che richiede anch'esso l'idrolisi dell'ATP.

Sappiamo molto poco della meccanica e dell'idrodinamica della traslocazione delle particelle, ma abbiamo trovato sulle vescicole strutture a forma di V (che interpretiamo come coppie di braccia laterali) che offrono un'elegante spiegazione in proposito. Mentre un braccio della V forma un ponte trasversale e procede nel ciclo che conduce al cambiamento conformazionale, l'altro braccio si attiva e, quando il primo finisce il ciclo e si stacca, si attacca a sua volta e via di seguito. Questo modo di procedere impedirebbe alla particella di venire spazzata via dal moto browniano che costantemente ne minaccia lo spostamento. Un simile meccanismo richiederebbe un numero di braccia laterali inferiore rispetto al modello che prevede solo braccia laterali liberamente dissociabili.

Come può il nostro modello spiegare il simultaneo spostamento degli organelli in entrambi i versi? Ho introdotto l'«ipotesi del contraccolpo», uno schema che prevede l'intervento di un solo braccio laterale per spiegare il trasporto delle particelle nei due versi. Ciascun braccio laterale potrebbe flettersi in modo tale che il colpo che parte in seguito ad attivazione faccia muovere un organello lungo una traiettoria relativamente lontana dalla superficie del microtubulo; poi il braccio laterale, durante il movimento di ritorno, si ritrarrebbe più vicino al microtubulo e gli organelli più grossi verrebbero così trasportati nel verso opposto. Le particelle che si spostano in versi opposti non si scontrerebbero mai perché occuperebbero piani diversi.

Il modello dell'attività intrinseca del microtubulo è solo ipotetico; c'è spazio per altre interpretazioni. Per esempio, Reese crede che esistano due diversi tipi di braccia laterali, ognuno responsabile del movimento in un solo verso. Sicuramente problemi come questi verranno risolti in futuro, dato che ora sappiamo visualizzare nel citoscheletro della cellula vivente il trasporto microtubulo-dipendente. Con il tempo chiariremo anche quel fenomeno che tanto ha destato il nostro interesse molti anni fa: il trasporto bidirezionale all'interno della cellula su quella che sembrava essere una via molecolare a senso unico.

I modelli per il trasporto, proposti dall'autore, presuppongono che sia lo scorrimento dei microtubuli, sia il movimento degli organelli siano controllati dalla stessa molecola: un «enzima generatore di forza» che viene sottoposto a ciclici cambiamenti di forma. Il sito di azione di questo enzima è nelle braccia laterali, che connettono i microtubuli agli organelli o interagiscono con il substrato. L'enzima generatore di forza può passare dallo stato attivo (*in rosso*) allo stato inattivo o di riposo (*in blu*). L'enzima nello stato inattivo si può attivare: (*a*) mentre è sul microtubulo; (*b*) mentre circola liberamente; (*c*) mentre è attaccato a un organello. Il primo modello è quello che si adatta meglio alla «ipotesi del contraccolpo», illustrata nella sezione trasversale. Esso suggerisce che i movimenti dell'enzima generatore di forza, in fase di attivazione e di inattivazione, avvengano su piani differenti, uno radiale e l'altro tangenziale; pertanto gli organelli si muoverebbero in versi opposti senza scontrarsi. L'ultimo modello (*c*) descrive il meccanismo che permette a un organello di spostarsi: il movimento viene coordinato da due o più braccia laterali vicine, che alternano lo stato attivo a quello inattivo in modo che l'organello sia in continuo contatto con il microtubulo. Poiché i microtubuli hanno una polarità intrinseca (indicata nella figura dai segni + e −), il legame delle braccia laterali può anche essere direzionale. In questa maniera un singolo microtubulo può scorrere in un solo verso che gli viene imposto dalla propria infrastruttura (*d*).

Sia le vescicole, sia i microtubuli possono presentare le braccia laterali, che prendono parte al trasporto microtubulo-dipendente. Quando è disponibile energia chimica sotto forma di ATP, le braccia aderiscono alla superficie delle vescicole (*a sinistra*); quando, invece, è presente l'analogo dell'ATP, l'AMP-PNP, le braccia laterali aderiscono ai microtubuli (*a destra*). Nelle microfotografie, le braccia laterali, ingrandite 210 000 volte, appaiono abbondanti sulle vescicole, ma scarse sui microtubuli, dove assumono l'aspetto di bastoncini rigidi posti tra le strutture globulari. Queste immagini sono state ottenute da George M. Langford della School of Medicine dell'Università del North Carolina.

Le pulsar più vecchie dell'universo

Queste stelle peculiari, dense e compatte, che ruotano a una velocità di varie centinaia di giri al secondo, devono forse le loro straordinarie caratteristiche al fatto di essere «risorte» dopo una precoce scomparsa

di Jacob Shaham

Un bel mattino verso la metà di novembre del 1982, un collega, David J. Helfand, astronomo della Columbia University, entrò nel mio ufficio per raccontarmi un fatto notevole. Era appena tornato da un viaggio di lavoro all'Osservatorio di Arecibo, a Puerto Rico, dove era venuto a sapere che Donald C. Backer dell'Università della California a Berkeley e collaboratori avevano scoperto una radiopulsar con un periodo di 1,558 millisecondi nella costellazione della Vulpecula. Questo valore, in caso di conferma, avrebbe reso la stella, contrassegnata dalla denominazione ufficiale 1937 + 214, la radiopulsar più veloce mai osservata.

La scoperta di Backer era una notizia davvero stupefacente. Le radiopulsar sono oggetti celesti dall'aspetto stellare che emettono molti tipi di radiazione elettromagnetica e particellare, tra i quali vi sono appunto le onde radio. Quando si punta un radiotelescopio verso una sorgente di questo genere, si ricevono le onde radio sotto forma di segnali periodici. La radiopulsar più rapida che si conoscesse prima della scoperta di 1937 + 214 era quella del Granchio, un'emittente con un periodo di 30 millisecondi immersa in una grande struttura sfumata simile a un batuffolo di cotone che prende il nome di Nebulosa del Granchio; adesso sembrava che ne esistesse una 20 volte più veloce.

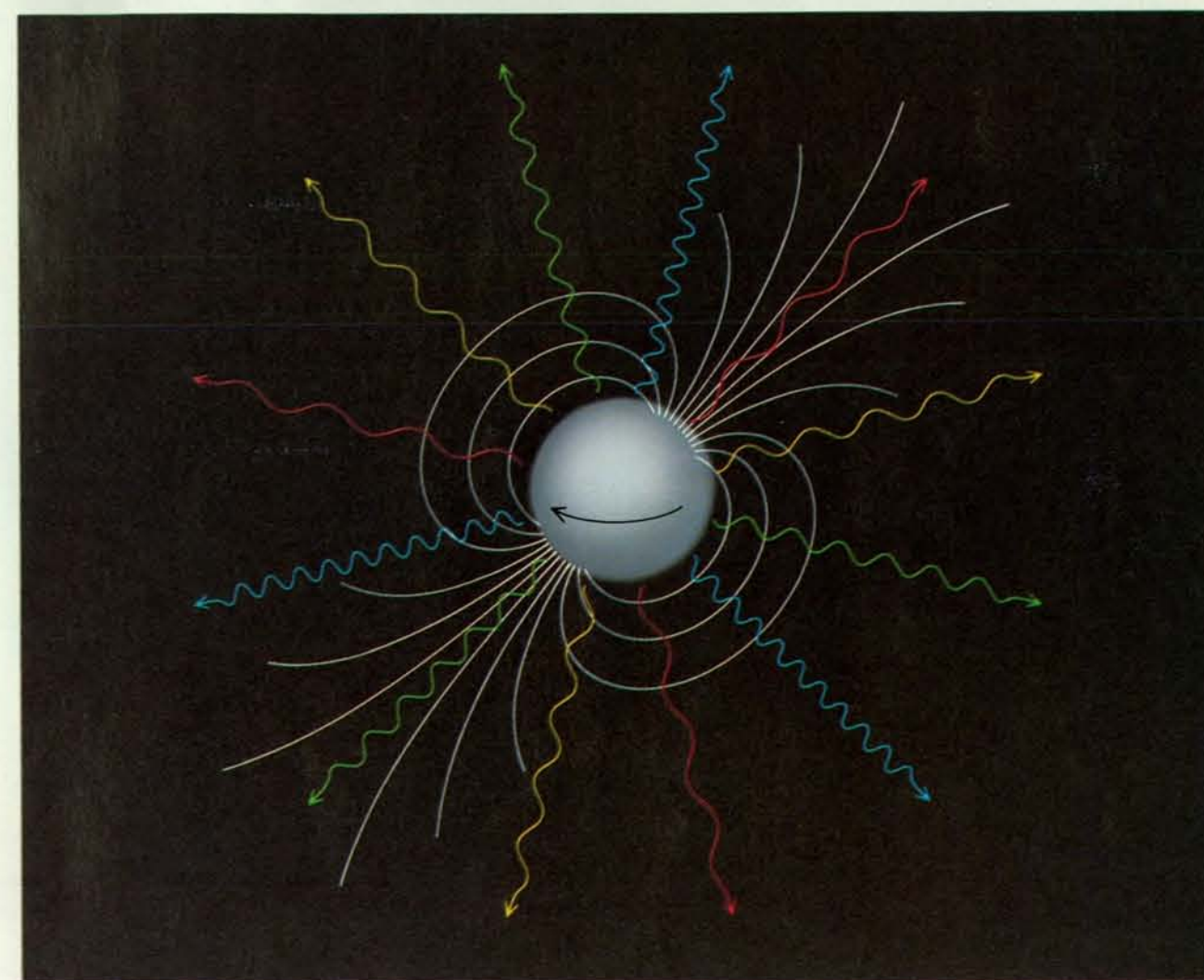
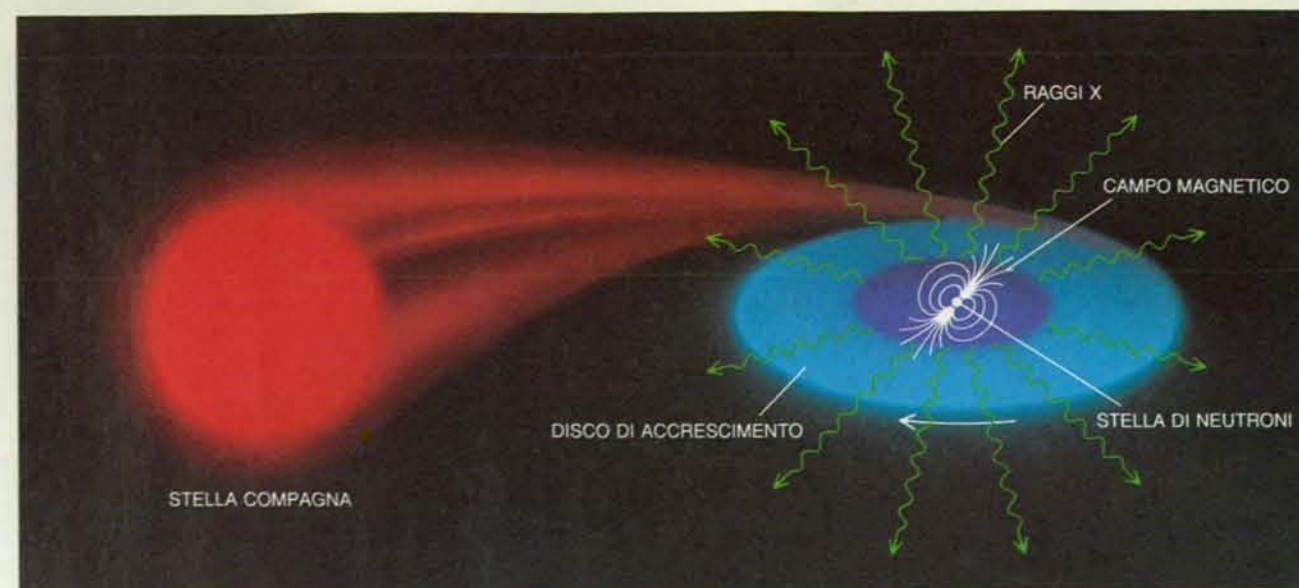
Dopo la scoperta della pulsar nella Vulpecula si è avuta notizia di altre due emittenti superveloci di questo tipo: nel 1983 è stata annunciata la scoperta di una pulsar con un periodo di 6,13 millisecondi (che è stata denominata 1953 + 29) e nel 1986 quella di una pulsar con un periodo di 5,362 millisecondi (1855 + 09), mentre si ventila la possibilità che ne sia stata osservata un'altra. A mano a mano che aumentano le indicazioni disponibili, appare sempre più verosimile che queste emittenti superveloci si possano spiegare solo annoverandole in una nuova categoria di pulsar.

Alcune delle caratteristiche generali di questo nuovo gruppo cominciano a profilarsi, anche se molti dei particolari restano oscuri: il dato più interessante è forse l'età attribuita a questi oggetti, che, a quanto si crede, sono molto vecchi. Le pulsar ordinarie sono relativamente giovani, avendo di norma un'età inferiore al milione di anni; la pulsar del Granchio, la più giovane che si conosca, ha soltanto 932 anni. Le pulsar superveloci, al contrario, sono molto vecchie, o almeno così si crede. Si tratta probabilmente del risultato di processi evolutivi che potrebbero aver avuto inizio addirittura un miliardo di anni fa, cioè un ventesimo dell'età dell'universo, ed essere destinati a protrarsi ancora per diversi miliardi di anni.

Le radiopulsar ordinarie sono affascinanti già di per sé: si ritiene che siano stelle di neutroni in rotazione, giganteschi «nuclei atomici» turbinanti che contengono qualcosa come 10^{57} protoni e neutroni. (Il nucleo dell'atomo di idrogeno, tanto per fare un confronto, comprende un solo protone e quello dell'atomo di ferro un totale di 56 fra protoni e neutroni.) Le stelle di neutroni si chiamano così perché contengono circa 20 volte più neutroni che protoni. Questo grande agglomerato di materia nucleare, che ha una massa circa pari a quella del Sole, è compresso in una sfera con un raggio dell'ordine della decina di chilometri. La densità della stella risulta di conseguenza enorme, leggermente superiore a quella della materia nucleare nel suo stato normale, la quale è già circa 10 bilioni di volte superiore a quella del piombo.

Le correnti di protoni e di elettroni che si muovono nello strato superficiale della stella generano un campo magnetico. Ruotando, la stella proietta un fascio di onde radio prodotto dall'effetto combinato del campo magnetico e della rotazione, il quale spazza periodicamente lo spazio tutto attorno come la luce di un faro. A ogni rotazione il fascio passa per la direzione in cui si trova la Terra, dando origine al segnale intermittente che i radiotelescopi rilevano. La pulsar del Granchio, per esempio, è una stella di massa pari a quella del Sole concentrata in una sfera delle dimensioni di una grande città, che ruota alla considerevole velocità di 33 giri al secondo (compiendo cioè una rotazione completa ogni 30 millisecondi).

Le proprietà di una stella di neutroni sono così peculiari che esaminandole ci si può divertire per parecchie ore. È facile calcolare, per esempio, la forza di gravità alla sua superficie, che risulta ben 100 miliardi di volte superiore a quella presente sulla superficie terrestre. In realtà, il campo gravitazionale di una stella di neutroni è così forte che qualunque oggetto attirato verso di essa si frantumerebbe prima dell'impatto: su una stella di neutroni gli oggetti non possono cadere, ma solo «piovere».

Quali sono i processi che potrebbero dare vita a una stella così straordinaria? Gli astrofisici ritengono che le stelle di neutroni si formino quando una stella ordinaria, con una massa parecchie volte superiore a quella del Sole, collassa violentemente nel fenomeno che prende il nome di esplosione di supernova. Le regioni interne della stella ordinaria implodono formando la stella di neutroni, mentre quelle esterne esplodono e danno origine a una nebulosa, una nube di gas e polveri, tutt'intorno a essa. Parte dell'energia dell'esplosione si libera sotto forma di luce visibile, che si manifesta in modo veramente spettacolare. Nel 1054 d.C., quando esplose la supernova che ha dato origine alla pulsar e alla nebulosa del Granchio, la luce emessa rimase visibile, di giorno, per quasi tre settimane e di notte per più di un anno. L'emissione di energia continua anche dopo la nascita della pulsar: la stella in

Si ritiene che la pulsar 1937 + 214, con un periodo di circa 1,5 millisecondi, abbia avuto un'origine diversa da quella di altre pulsar. Si pensa infatti che si tratti di una densa stella di neutroni «risorta». Forse, decine di milioni di anni fa o più la stella faceva parte di una binaria X (*in alto*), ossia di un sistema costituito da una stella di neutroni in rotazione su se stessa, attorno alla quale orbita una compagna. Secondo questa ipotesi, mentre le due stelle ruotavano l'una intorno all'altra, quella di neutroni accumulava lentamente materia proveniente dalla compagna. L'impatto della materia in arrivo sulla superficie della stella avrebbe causato l'emissione di raggi X e l'accelerazione della sua rotazione su se stessa. Quando alla fine il processo di accrescimento ebbe fine con l'esaurimento totale della compagna, la stella di neutroni presumibilmente ruotava abbastanza rapidamente da essere una radiopulsar (*in basso*). Nell'illustrazione le linee ondulate rosse indicano le onde radio, quelle gialle la luce visibile, quelle verdi i raggi X e quelle blu i raggi gamma. Non è chiaro perché la compagna sia scomparsa.

rotazione emette particelle altamente energetiche che vengono assorbite dalla nebulosa circostante e la fanno risplendere. Via via che emette particelle, la pulsar perde energia di rotazione e rallenta. Vale la regola empirica secondo la quale a pulsar più vecchie corrispondono velocità di rotazione inferiori.

Ci si potrebbe attendere, quindi, che le pulsar superveloci scoperte negli ultimi tempi siano molto giovani, sicuramente più di quella del Granchio, che ha 932 anni. A cavallo tra il 1982 e il 1983, però, M. Ali Alpar dell'Istituto di ricerca per le scienze di base della Turchia, Andrew F. Cheng della Johns Hopkins University e Malvin A. Ruderman della Columbia University si sono uniti a chi scrive nell'analisi dei dati relativi alla pulsar superveloce 1937 + 214 raccolti da Backer e colleghi (S. R. Kulkarni e Carl E. Heiles di Berkeley, M. M. Davis dell'Osservatorio di Arecibo e W. M. Goss del Kapteyn Laboratory nei Paesi Bassi). Sia i nostri calcoli, sia quelli eseguiti da altri gruppi hanno indicato che la velocità di rotazione è così elevata da essere inferiore solo di un fattore due circa alle condizioni limite oltre le quali la pulsar si smembrerebbe per effetto della forza centrifuga. Appariva dunque chiaro che quando era nata, la stella non poteva ruotare molto più rapidamente e che pertanto, da allora, non poteva aver rallentato molto.

Alla fine è stata però una considerazione di ordine statistico a portare alla conclusione che la pulsar superveloce, pur avendo una velocità di rotazione molto elevata, è probabilmente un oggetto assai antico. In altre parole, sembra che questa stella si sia formata durante un processo evolutivo lungo e non nel corso dell'esplosione relativamente rapida che caratterizza la nascita di una supernova.

La considerazione statistica riguarda l'entità del rallentamento delle radiopulsar. Esprimendosi in parole semplici, si può affermare che quanto più breve è il periodo di una pulsar, tanto più sensibile è il rallentamento e che, inversamente, pulsar con periodo lungo rallentano in misura minore. Di conseguenza dovrebbero esistere più pulsar lente che pulsar veloci. A meno di non trovarsi di fronte a un caso anomalo, l'esistenza della pulsar 1937 + 214 con un periodo di 1,5 millisecondi comporta che, nell'ipotesi che le pulsar nascano a un tasso abbastanza costante nel tempo, oggi dovrebbero esistere molte pulsar con un periodo lungo, compreso tra 10 e 30 millisecondi. Eppure, nonostante ricerche abbastanza approfondite condotte sulla sfera celeste, non si è trovata alcuna pulsar di questo tipo. Inoltre, pur avendo un periodo così breve, 1937 + 214 non sta affatto rallentando. Perché continua a ruotare così velocemente?

Per spiegare il disaccordo fra teoria e osservazioni abbiamo sfruttato la circostanza che la velocità di rallentamento di una pulsar dipende anche dall'intensità del campo magnetico alla superficie della stella. Quanto più intenso è il campo, tanto maggiore è l'energia che la pulsar irradia e quindi più rapido il suo rallentamento. Una tipica stella di neutroni possiede un campo magnetico di circa un bilione di gauss, pari a circa 100 milioni di volte quello di un magnete artificiale molto potente. Noi abbiamo dimostrato che se una pulsar con periodo di 1,5 millisecondi avesse un campo magnetico 1000 volte inferiore (pari cioè a circa un miliardo di gauss), potrebbe plausibilmente occorrerle un lasso di tempo paragonabile all'età dell'universo per rallentare fino a raggiungere un periodo di 10 millisecondi. Questa circostanza potrebbe spiegare perché non siano state ancora osservate pulsar con periodo compreso tra 10 e 30 millisecondi: non avrebbero avuto tempo bastante per rallentare fino a questi valori.

Qualche semplice calcolo indica inoltre che, se il campo magnetico di una pulsar è relativamente debole, la stella può anche «spegnersi» e smettere di irradiare energia prima di aver subito un rallentamento significativo. Una pulsar ordinaria, dotata di un campo magnetico di un bilione di gauss, si spegne quando la velocità di rotazione scende al di sotto di un giro al secondo circa. Una pulsar con un campo magnetico di 100 milioni di gauss, invece, si spegnerebbe già a una velocità di rotazione inferiore a un giro ogni 10 millisecondi. In altre parole, una pulsar con un campo magnetico debole, se anche fosse riuscita in qualche modo a rallentare fino a quella velocità, non risulterebbe comunque osservabile.

Tutte queste considerazioni portano a un risultato degno di nota: l'esistenza stessa della pulsar 1937 + 214 indica che deve esistere una nuova categoria di radiopulsar, profondamente diversa da quella delle pulsar ordinarie, costituita da stelle con un debole campo magnetico superficiale.

La conclusione che esiste una nuova categoria di radiopulsar dotate di campo magnetico superficiale debole ci ha fatto trarre un gran sospiro di sollievo. L'osservazione della radiazione proveniente da 1937 + 214 e dalle sue vicinanze indica che la pulsar libera relativamente poca energia. Eppure se una pulsar ordinaria, cioè con campo magnetico intenso, ruotasse a velocità pari a quella di 1937 + 214, il suo «tempo caratteristico di rallentamento» (il tempo necessario perché il periodo raddoppi) sarebbe di pochi anni e si potrebbe osservare un'enorme emissione di energia: la stella sarebbe vari milioni di volte più luminosa della pulsar del Granchio. Abbiamo invece calcolato che, essendo una pulsar a campo debole, 1937 + 214 ha un tempo caratteristico di rallentamento di quasi un miliardo di anni e un'emissione di energia inferiore di un fattore pari a circa 100 milioni, un valore coerente con le osservazioni. La nostra teoria riesce pertanto a spiegare quella che altrimenti risulterebbe una discrepanza di grande entità.

Una stella di neutroni - qui raffigurata in sezione secondo tre diversi modelli teorici - è costituita da neutroni e protoni serrati così strettamente che la sua densità supera persino quella della materia nucleare normale, che è già di per sé circa 10 bilioni di volte più densa del piombo. La stella qui raffigurata ha una massa pari a 1,4 volte quella del Sole, eppure il suo raggio è compreso tra 10 e 16 chilometri appena, paragonabile cioè a quello di una grande città.

A questo punto ci siamo trovati di fronte a una nuova svolta nell'analisi. Come si crea una pulsar con un campo magnetico debole? Il caso volle che due settimane prima di venire a conoscenza dell'elevata velocità di rotazione di 1937 + 214 avessimo occasione di notare un nuovo studio di Roger D. Blandford del California Institute of Technology, James H. Applegate della Columbia University e Lars Hearnquist di Berkeley. In questo lavoro essi dimostravano che se una pulsar si forma ad alta temperatura (sopra i 100 milioni di kelvin circa, una temperatura che in realtà è inferiore a quella caratteristica di un'esplosione di supernova), acquista automaticamente un forte campo magnetico. La conseguenza era evidente: la pulsar 1937 + 214 doveva essersi formata «a freddo», a una temperatura inferiore a 100 milioni di kelvin, e pertanto non poteva essere stata prodotta in un'esplosione di supernova. Infatti, dal momento che una nascita «a caldo» comporta un campo magnetico intenso e una velocità di rallentamento elevata, non si sarebbe potuta formare una pulsar dotata nel contempo di un periodo breve di 1,5 millisecondi e di un campo magnetico debole.

L'unico sistema che si conosca per imprimere a una stella di neutroni fredda un'elevata velocità di rotazione è il lento accumularsi di materiale in rotazione. Il fenomeno alla base è ben noto a chiunque abbia osservato un pattinatore che si mette a girare velocemente su se stesso: all'inizio gira lentamente a braccia allargate, poi le stringe al tronco, facendo così aumentare la propria velocità di rotazione. Perché si verifichi un fenomeno analogo in una stella di neutroni fredda, il suo tasso di accrescimento deve essere lento, altrimenti la stella si surriscalderebbe e il campo magnetico balzerebbe a valori troppo alti. Di conseguenza la pulsar 1937 + 214 deve essere rimasta in formazione per molto tempo, di certo per almeno 10 milioni di anni e probabilmente per un periodo dell'ordine del miliardo di anni.

L'accrescimento lento potrebbe verificarsi in due modi. Secondo un primo scenario ipotetico su una cosiddetta nana bianca (una stella avente la stessa massa di una stella di neutroni, ma un raggio 1000 volte maggiore) si deposita materia proveniente da una compagna che le orbita attorno. Via via che la materia si accumula, la nana bianca collassa trasformandosi in una stella di neutroni. Se il collasso fosse abbastanza lento, la stella di neutroni potrebbe accelerare progressivamente la propria rotazione e diventare una pulsar con un periodo di qualche millisecondo. Accumulando ancora materia dalla compagna, la stella potrebbe poi aumentare la propria velocità di rotazione fino a raggiungere un periodo di rotazione all'equilibrio.

In alternativa, la stella di neutroni potrebbe essersi formata nel solito modo, cioè in un'esplosione di supernova, e comportarsi come una pulsar normale. In tal caso, dopo circa 10 milioni di anni, le correnti elettriche presenti nel suo strato più superficiale svanirebbero quasi completamente, lasciando un debole campo magnetico. Se la pulsar «morta» possedesse una compagna che le gira attorno su un'orbita stretta, su di essa potrebbe accumularsi il materiale necessario a farle assumere una velocità di rotazione elevata e quindi a risorgere, o riciclarsi. In questa situazione, come nella precedente, la compagna dovrebbe essersi sacrificata interamente nel processo, dal momento che nelle vicinanze di 1937 + 214 non si osservano stelle.

Alla base di entrambi gli scenari si pensa che possa esserci un unico meccanismo il quale assicura che il processo di accrescimento non si svolga troppo velocemente. Quando gli elettroni, gli atomi e le molecole in arrivo colpiscono la superficie della stella parte della loro energia si trasforma in luce visibile e in altri tipi di radiazione elettromagnetica, la quale si muove verso l'esterno colpendo altri elettroni, atomi e molecole in

Considerazioni statistiche indicano che la pulsar 1937 + 214 deve possedere un campo magnetico molto più debole di quello di una pulsar comune. Il grafico, elaborato sulla base dell'ipotesi che la pulsar abbia un campo magnetico di circa un bilione di gauss, ripercorre l'evoluzione di una famiglia di pulsar con periodo di qualche millisecondo (*curve in nero*). La spaziatura costante delle curve riflette l'ipotesi che le pulsar nascano a un tasso più o meno costante nel tempo. Durante l'invecchiamento, la velocità di rotazione della stella diminuisce e quindi il periodo si allunga. In generale il periodo di una pulsar varia a una velocità circa inversamente proporzionale al valore del periodo stesso. In altre parole, quanto più corto è il periodo di una pulsar, tanto più veloce è il rallentamento, mentre a periodi più lunghi corrisponde un rallentamento meno rapido. Di conseguenza l'esistenza della pulsar con periodo di pochi millisecondi (*gruppo inferiore di linee tratteggiate*) comporta l'esistenza di parecchie pulsar con periodo di circa 20 millisecondi (*linee tratteggiate in alto*). Tuttavia non si è scoperta alcuna pulsar di quest'ultimo tipo. È possibile spiegare questa discrepanza postulando che le pulsar con periodo di qualche millisecondo abbiano un campo magnetico debole, dell'ordine del miliardo di gauss. Dal momento che hanno un campo magnetico più debole di quello delle pulsar ordinarie, le pulsar di brevissimo periodo irradiano meno energia e potrebbero impiegare un tempo paragonabile all'età dell'universo per rallentare fino a un periodo di 10 millisecondi. Può darsi addirittura che, quando raggiungono questo periodo, si siano esaurite, cioè abbiano smesso di irradiare.

avvicinamento. Poiché possiede una certa quantità di moto, tale radiazione esercita su questa materia una forza che è diretta verso l'esterno e che si oppone quindi all'attrazione gravitazionale centripeta della stella.

Se il tasso di accrescimento diventa abbastanza elevato (6×10^{17} grammi al secondo, cioè 10^{-8} masse solari all'anno), la forza esercitata dalla radiazione è tanto intensa da equilibrare l'attrazione gravitazionale e il tasso di accrescimento non può aumentare ulteriormente. Il tasso al quale le due forze si equilibrano prende il nome di «limite di Eddington» dal nome dell'astronomo inglese Sir Arthur Eddington. In parole semplici, l'esistenza del limite di Eddington significa che qualsiasi processo di accrescimento su una stella di neutroni possiede un «termostato» incorporato il quale garantisce che il tasso di accrescimento non sia mai troppo alto; se la materia disponibile per l'accrescimento è sufficiente, però, esso non sarà neppure mai troppo basso. Sembra quindi che questo termostato sia fatto su misura per la formazione di pulsar aventi un periodo di qualche millisecondo, perché limitare il tasso di accrescimento significa limitare le temperature stellari a 100 milioni di gradi o meno.

Si conosce già un processo dipendente dall'accrescimento che svolge un ruolo importante nella formazione delle «sorgenti X compatte», piccole stelle di neutroni in rotazione che emettono raggi X: in una di queste classi di sorgenti, quella delle binarie X di piccola massa, la stella di neutroni accumula massa strappata a una compagna leggera, in un processo che può continuare per 100 milioni di anni o più.

Le radiopulsar si esauriscono quando il campo magnetico superficiale diventa troppo debole o quando il periodo di rotazione diventa troppo lungo. Il grafico mostra il confine tra le radiopulsar attive e quelle «morte» al variare del campo e del periodo. Si noti che, poiché una stella di neutroni si smembra se il periodo è inferiore al millisecondo, il campo superficiale di una pulsar attiva deve essere di almeno 100 000 gauss, circa 10 volte quello di un potente magnete artificiale.

Una particella carica che si muova in direzione perpendicolare a un campo magnetico viene deviata dalla forza di Lorentz così da percorrere una traiettoria di forma tendenzialmente circolare. Finché il campo è debole e la particella veloce, la deviazione risulta trascurabile (*in basso*). Via via che il rapporto tra intensità del campo e velocità cresce, le deviazioni diventano sempre maggiori, finché oltre un certo valore il campo costringe la particella a tornare indietro (*in alto*).

Il fatto che le binarie X di piccola massa presentino una così lunga fase di accrescimento ha sollevato un problema interessante: è possibile che questi oggetti siano i progenitori delle pulsar con periodo di qualche millisecondo? Noi siamo giunti alla conclusione che deve essere senz'altro così: sia lo scenario con la nana bianca che collassa, sia quello con la pulsar morta richiedono un sistema stellare binario longevo, con un processo di accrescimento in atto, e gli unici oggetti che si conoscano in grado di svolgere questo ruolo sono le binarie X di piccola massa.

Considerazioni statistiche sembrano corroborare l'idea che le binarie X di piccola massa siano le antenate delle pulsar con periodo di qualche millisecondo. Queste pulsar devono vivere molto a lungo prima di estinguersi; si ammetta ai fini di questo calcolo che abbiano una vita di 20 miliardi di anni, pari all'età dell'universo. Si stima che la vita delle binarie X di piccola massa sia relativamente breve: all'incirca 100 milioni di anni. Se le cose stanno così, il rapporto tra pulsar con periodo di qualche millisecondo e binarie X di piccola massa dovrebbe essere, grosso modo, quello tra 20 miliardi e 100 milioni, ossia 200 a 1.

La conclusione è coerente con le osservazioni. Come si è detto più sopra, finora sono state scoperte in totale tre pulsar con periodo di qualche millisecondo; certamente se ne scopriranno altre, e poiché esse non si vedono dalla Terra se non sono abbastanza vicine, è ragionevole stimare che diverse migliaia di pulsar di brevissimo periodo si nascondano non viste nella Galassia. Sulla base di queste considerazioni ci si può quindi aspettare di vedere in tutta la Galassia poche decine di binarie X di piccola massa, proprio come accade nella realtà.

Le considerazioni statistiche sono di per sé rassicuranti, ma per noi è forse più importante il fatto di essere riusciti a elaborare un modello che prevede esattamente come varia la velocità di rotazione della pulsar 1937 + 214 usando solo alcuni principi fisici di base. Secondo la nostra teoria, 1937 + 214 è una radiopulsar risorta, che ha avuto origine come membro di un sistema binario e ha aumentato la propria velocità di rotazione per accrescimento. Quando questo è terminato la stella aveva una velocità di rotazione abbastanza elevata da diventare una radiopulsar nonostante il campo magnetico relativamente debole. Abbiamo calcolato che il periodo dell'oggetto dovrebbe crescere in ragione di 10^{-20} o 10^{-19} secondi per secondo, un valore inferiore di molti ordini di grandezza a quello che ci si sarebbe aspettato per qualunque pulsar veloce conosciuta prima del 1982. La nostra previsione è stata confermata: il valore misurato per l'incremento di crescita del periodo è pari a $1,24 \times 10^{-19}$ secondi per secondo, con un errore di $0,25 \times 10^{-19}$ secondi per secondo.

L'accumulo di materiale su una stella di neutroni ha luogo principalmente in prossimità dei poli perché è là che le linee di forza del campo magnetico iniziano e finiscono. Le particelle cariche che si muovono parallelamente rispetto al campo magnetico non vengono deviate e quindi tendono ad accumularsi ai poli. Le particelle che si muovono in direzione perpendicolare a quella delle linee di forza vengono invece deviate, ma le collisioni con le altre particelle in arrivo impediscono loro di tornare indietro. La distanza alla quale le particelle tendono ad arrestarsi definisce il raggio della magnetosfera; nella regione al suo interno, la magnetosfera appunto, le particelle si muovono prevalentemente lungo le linee di forza e ruotano insieme alla stella.

Ma la soluzione completa del mistero di 1937 + 214 non è ancora a portata di mano. In particolare, non è chiara una cosa: se l'aumento della velocità di rotazione della pulsar si deve al suo accrescimento, dove si trova la compagna? Questa domanda è diventata più pressante nel 1983, con la scoperta (dovuta a Valentin Boriakoff della Cornell University, R. Buccheri del Consiglio nazionale delle ricerche a Palermo e F. Fauci dell'Università di Palermo) della pulsar 1953 + 29, la quale invece possiede una compagna.

Poco tempo dopo questa scoperta, insieme a Helfand e a Ruderman ho incominciato a riflettere sul motivo per il quale 1953 + 29 fa parte di un sistema binario mentre 1937 + 214 è isolata. Per capirlo abbiamo sfruttato la circostanza che esistono due tipi di binarie X di piccola massa, quelle brillanti e quelle deboli. Le prime si possono interpretare solo come sistemi comprendenti una stella «normale», che si sarebbe espansa fino a trasformarsi in una gigante rossa, per perdere poi l'involucro esterno di materia in favore di una stella di neutroni compagna. Le binarie deboli, invece, si formerebbero quando due stelle si avvicinano, perdendo una parte dell'energia cinetica del moto orbitale ed emettendo radiazione gravitazionale.

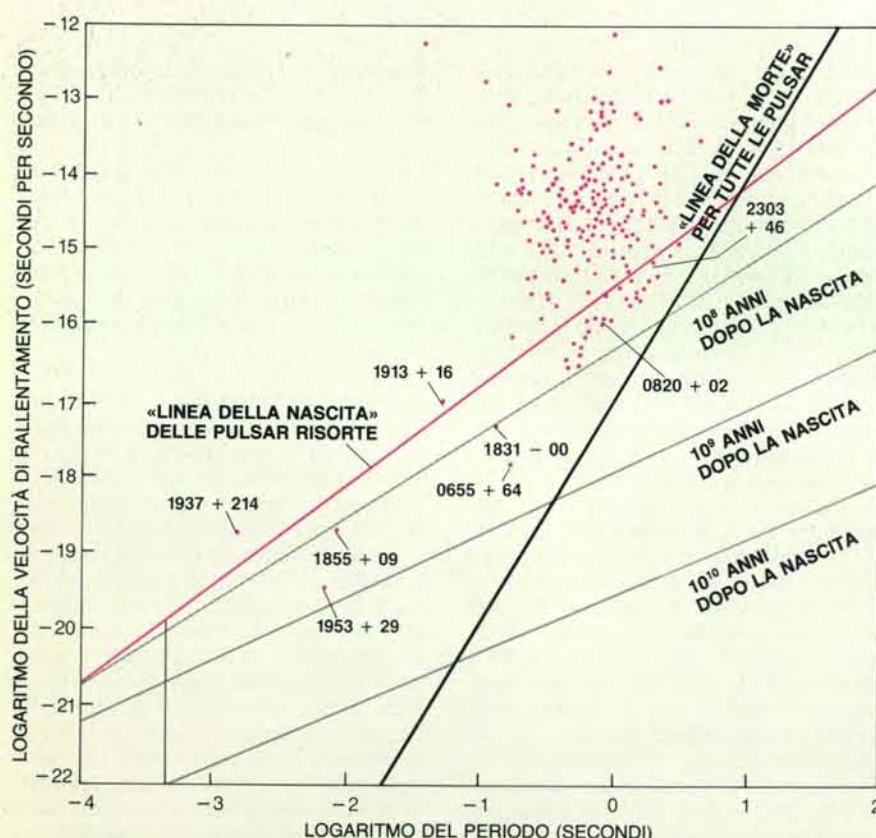
Negli ultimi cinque anni circa, vari ricercatori hanno dimostrato che l'evoluzione di una binaria X brillante di piccola massa può dare origine solo a un sistema binario, che è allora costituito da una pulsar e dal nucleo residuo della stella «normale». (Tra gli studiosi che sono giunti a questa conclusione vi sono Paul C. Joss e Saul A. Rappaport del Massachusetts Institute of Technology, Edward P. J. van den Heuvel e G. J. Savonije dell'Istituto astronomico di Amsterdam, Ronald E. Taam della Northwestern University, Ronald F. Webbink dell'Università dell'Illinois a Urbana-Champaign.) Perciò abbiamo ipotizzato che 1953 + 29 abbia avuto inizio come binaria X brillante di piccola massa e 1937 + 214 sia stata in origine una sorgente di tipo analogo, ma debole.

La situazione si è fatta più complessa nel 1986 con la scoperta (da parte di D. J. Segelstein, L. A. Rawley, D. R. Stinebring, A. S. Fruchter e Joseph H. Taylor della Princeton University) della pulsar 1855 + 09, avente un periodo di 5,362 millisecondi e anch'essa parte di un sistema binario. Sembra che nelle linee generali la storia di 1855 + 09 e quella di 1953 + 29 siano piuttosto simili sia tra di loro, sia a quella di cinque altre pulsar binarie le quali, pur non essendo superveloci, hanno un periodo piuttosto breve, pari ad alcune decine di millisecondi. 1937 + 214, priva di compagna com'è, comincia pertanto a diventare un caso davvero peculiare. Sono stati proposti diversi scenari per spiegare come la compagna possa essere sparita e in almeno uno di questi scenari si ipotizza che la pulsar sia relativamente giovane, con un'età di circa un milione di anni. Fino a oggi, però, nessuna spiegazione riscuote pieno successo e la ricerca di un'ipotesi del tutto soddisfacente continua.

Forse però l'anello mancante critico nella nostra storia era e resta un altro. Nessuna binaria X di piccola massa finora osservata mostra di annoverare tra i suoi membri una stella di neutroni superveloce, mentre ci si aspetterebbe che, per dar vita a una pulsar superveloce, una binaria X di piccola massa debba contenere una stella di neutroni in rotazione con un periodo di pochi millisecondi. Perché non si è ancora osservata alcuna periodicità di questo tipo?

Varie spiegazioni si presentano spontaneamente all'attenzione. La più plausibile parte dal fatto che qualsiasi stella di neutroni in una binaria X di piccola massa sia probabilmente assai vecchia, e possieda di conseguenza un campo magnetico che è solo il residuo di uno precedente, più intenso; questa condizione è necessaria perché, al termine del periodo di accrescimento, si formi una pulsar con un periodo di qualche millisecondo. David S. Eichler e Zhengzhi Wang dell'Università del Maryland a College Park hanno dimostrato che l'intensità del campo magnetico residuo può variare notevolmente da un punto all'altro della superficie. Perciò, mentre una stella di neutroni ordinaria ha due «punti caldi» del campo magnetico che emettono raggi X (due regioni polari dalle quali provengono i raggi X), può darsi che una stella di neutroni vecchia possa avere un certo numero di questi punti sulla sua superficie.

Se la stella ruota e la si osserva con un telescopio per raggi X, i punti caldi si manifesteranno come variazioni di intensità del flusso di raggi X. Queste va-

Il grafico mostra la variazione del periodo di equilibrio e del raggio della magnetosfera in funzione del tasso di accrescimento per varie intensità del campo magnetico. Il periodo di equilibrio è quello in corrispondenza del quale la materia che si trova a una distanza pari al raggio della magnetosfera ruota insieme alla stella. I calcoli riguardano una stella di 1,4 masse solari.

Nel grafico della velocità di rallentamento osservata in funzione del periodo, le pulsar ordinarie, nate da un'esplosione di supernova, si raggruppano nella regione in alto a destra. Col tempo, esse scivolano in basso e verso destra fino a raggiungere la «linea della morte». Sono state scoperte sei pulsar al di fuori del gruppo principale, che, insieme ad altre due che si trovano all'interno del gruppo, appartengono a una nuova categoria di pulsar la cui velocità di rotazione deve essere aumentata grazie a un processo di accrescimento lento. Tutti i membri della nuova classe, che si trovano naturalmente al di sopra della linea della morte, hanno una compagna con la quale formano un sistema binario, eccetto 1937 + 214. Le curve in colore e in grigio indicano fasi della vita di pulsar accelerate in base a un tasso di accrescimento di 10^{-9} masse solari all'anno. La linea verticale indica il periodo di rotazione minimo, al di sotto del quale la stella si spacca.

riazioni non saranno però molto intense né molto nette; avranno l'aspetto di sfavillii irregolari a frequenze molto più elevate. Anche se si ripeterà a ogni rotazione della stella, lo schema di variazione dell'intensità risulterà così offuscato che il periodo, e quindi la stessa stella in rotazione, potrebbero risultare impossibili da rilevare contro il fondo di rumore X proveniente dalla sorgente.

Esistono altri effetti che potrebbero rendere difficile individuare la presunta rotazione della stella di neutroni facente parte di una binaria X di piccola massa: la materia proveniente dalla compagna che si accumula sulla stella potrebbe diffondere i raggi X emessi dai punti caldi riducendo l'entità della modulazione del segnale. Inoltre la stella di neutroni funge probabilmente da «lente gravitazionale», deflette cioè i raggi X che ne escono, riducendo ancor più l'entità della modulazione.

Queste considerazioni sono valide, tuttavia sarebbe bello trovare effettivamente una stella di neutroni in rotazione in una binaria X di piccola massa.

Nel 1983 la European Space Agency ha lanciato in orbita attorno alla Terra il satellite EXOSAT, un osservatorio orbitante per raggi X. Uno degli esperimenti a bordo del satellite era volto alla ricerca sistematica di periodi molto brevi nell'emissione X delle binarie X di piccola massa. Non è stato scoperto alcun periodo ben definito, ma sono stati raccolti alcuni segnali quasi periodici, sfavillii intermittenti della radiazione, nella sorgente galattica GX5-1. L'intervallo fra queste «eruzioni» (*burst*) variava tra 25 e 50 millisecondi.

Questi fenomeni quasi periodici manifestano alcune proprietà insolite, come hanno riferito Michel van der Klis della European Space Agency e colleghi (F. Jansen del Laboratorio per la ricerca spaziale e J. van Paradijs e M. Sztajno dell'Istituto astronomico di Amsterdam, J. Trümper del Max Planck Institut für Physik und Astrophysik di Monaco e W.H.G. Lewin del MIT e del Max Planck Institut) nella circolare numero 4043 dell'Unione astronomica internazionale del 13 marzo 1985 e, in seguito, su «Nature». Innanzitutto, le quasiperiodicità non sono costanti ma presentano, in realtà, una variabilità notevole: il segnale quasiperiodico è più forte nei quasiperiodi più lunghi osservati e cala sistematicamente nei quasiperiodi più brevi. La caratteristica più curiosa del fenomeno è che la luminosità totale di GX5-1 è in stretto rapporto con la lunghezza del quasiperiodo: quanto più breve è questo, tanto più elevata è quella.

Parecchie settimane prima che uscisse la circolare dell'Unione astronomica internazionale, Alpar, che a quell'epoca si trovava all'Università dell'Illinois, mi telefonò per parlarmene. Un suo amico in Europa aveva partecipato a un seminario su quei risultati, e la notizia si era sparsa. Per quanto fossero curiosi, i risultati sembravano genuini: rendendoci conto che contenevano un indizio importante per risolvere il problema delle pulsar superveloci, ci mettemmo quindi a lavorare freneticamente per elaborare un modello che fosse in grado di spiegare le osservazioni.

Nel nostro modello i segnali quasiperiodici provengono dalla materia in orbita intorno alla stella di neutroni che tenta di «agganciarsi» al campo magnetico, il quale naturalmente ruota con la stella. Per quanto riguarda le «eruzioni», esse si verificano con una frequenza pari alla differenza tra la frequenza orbitale della materia di accrescimento e quella di rotazione della stella. L'intervallo di variabilità delle frequenze orbitali porta a un comportamento quasiperiodico invece che a uno ben definito. Secondo la nostra analisi, GX5-1 deve ruotare con un periodo compreso tra sei e 10 millisecondi circa e avere un campo magnetico alla superficie di circa un miliardo di gauss. A questo punto il nostro entusiasmo è stato davvero notevole, perché entrambi questi valori sono notevolmente vicini a quelli rilevati per la pulsar binaria 1953 + 29.

Finalmente avevamo un'indicazione chiara dell'esistenza di una periodicità dell'ordine di qualche millisecondo all'interno di una sorgente di raggi X di piccola massa! Sembrava proprio che le tessere di tutto il mosaico della formazione delle pulsar con periodo di qualche millisecondo alla fine si componessero.

La scoperta della sorgente X quasi periodica GX5-1, che oggi viene chiamata pressoché universalmente «oscillatore quasiperiodico», ha proiettato l'astronomia X in un turbine di avvenimenti. Fino a oggi sono stati scoperti più di 10 oscillatori di questo tipo, sette dei quali presentano un periodo breve (da 25 a 250 millisecondi). Questi sette oggetti sono tutti associati a binarie X di piccola massa e alta luminosità. Il comportamento da manuale di GX5-1 ha lasciato il campo, negli altri oscillatori, a una svariata gamma di fenomeni. Tutte queste complicazioni scateneranno senza dubbio negli anni a venire ulteriori discussioni ed elaborazioni di modelli per gli oscillatori quasiperiodici e le pulsar con periodo di qualche millisecondo.

Resta ancora molto da fare, ma comincia a profilarsi il quadro affascinante della resurrezione delle pulsar superveloci e longeve dalle antiche stelle di neutroni. Questa immagine ci offre anche un nuovo scorcio su alcuni dei più vecchi sistemi stellari della Galassia. Le pulsar hanno già portato a intuizioni nuove e brillanti nelle aree più disparate, dalla fisica delle particelle a quella nucleare e a quella dello stato solido, da quella dei plasmi a quella dei superfluidi, fino all'elettromagnetismo e alla relatività generale: oggi le pulsar più antiche illuminano la strada dell'evoluzione della Galassia.

I cromosomi «arlecchino» sono il risultato del reciproco scambio di materiale genetico tra cromatidi fratelli (cioè tra i due filamenti identici che costituiscono un cromosoma durante la divisione cellulare). Le immagini sono state ottenute da Sheldon Wolff, dell'Università della California a San Francisco, con una tecnica mediante la quale i cromatidi sono stati alterati chimicamente in modo che uno emettesse una fluorescenza più intensa del cromatidio fratello, quando veniva colorato con una sostanza fluorescente (*in alto*). Un agente che altera il DNA, la molecola costitutiva dei cromosomi, ha dato luogo a molteplici scambi (*in basso*) che sono analoghi alla ricombinazione genetica perché i cromatidi rimangono completi, ossia non acquisiscono né perdono informazioni genetiche. Diversamente dalla ricombinazione, che produce nuovi assortimenti di caratteri in quanto i due cromosomi sono diversi, lo scambio tra cromatidi fratelli non ha alcun effetto genetico.

La ricombinazione genetica

Prima di riprodursi, un organismo spesso rimescola il proprio materiale genetico: i suoi cromosomi possono scambiarsi segmenti in un processo di ricombinazione del quale si stanno svelando i meccanismi molecolari

di Franklin W. Stahl

Per riprodursi con successo un organismo deve trasmettere alla progenie una copia fedele della propria informazione genetica. Stranamente, però, fa anche parte dello schema fondamentale della riproduzione che un'aliquota di quella informazione sia rimescolata tra i vari depositari prima di essere trasmessa alla generazione successiva. I depositari sono i cromosomi: lunghe molecole di DNA a doppio filamento nelle quali l'informazione genetica è codificata in una sequenza di unità chimiche appaiate, le basi nucleotidiche. Il rimescolamento è noto con il nome di ricombinazione genetica.

Molto prima del 1953, anno in cui James Watson e Francis H. C. Crick descrissero la base molecolare dell'eredità, cioè la doppia elica di DNA, si sapeva che i singoli geni, le unità dell'informazione genetica, possono cambiare i propri «compagni di viaggio». Due caratteri associati a un singolo cromosoma in un individuo apparivano talvolta associati a cromosomi diversi nella progenie di quell'individuo. In effetti, la misurazione della frequenza con la quale due geni qualunque sono trasmessi contemporaneamente su un cromosoma è diventata un sistema standard per valutare la loro separazione fisica: quanto più i geni sono vicini, tanto più è probabile che essi vengano trasmessi insieme.

Tuttavia, senza le basi teoriche fornite dalla scoperta di Watson e Crick, non ci si potrebbe neppure chiedere in che modo avvenga esattamente il rimescolamento. Ora, a più di tre decenni di distanza, studi effettuati sui più semplici tra gli organismi, ossia i virus e i batteri, hanno chiarito almeno alcuni meccanismi molecolari, raffinati e complessi, della ricombinazione.

Si può pensare che analoghi meccanismi operino in altri esseri viventi, tra i quali la specie umana: la ricombinazione genetica, come il DNA su cui essa opera, è quasi universale. Negli organismi complessi essa ha luogo in genere durante la meiosi, il processo di divisione cellulare che produce le cellule germinali (uova e spermatozoi).

Nei precursori delle cellule germinali, come nelle cellule degli altri tessuti dell'organismo, sono presenti due serie di cromosomi: ogni cromosoma ha, infatti, una «controparte», un omologo, che porta un corredo di informazioni genetiche pressoché identico. Una serie di cromosomi deriva dalla cellula germinale materna, la cellula uovo, e l'altra dalla cellula germinale paterna, lo spermatozoo. Durante la meiosi, i cromosomi omologhi si appaiano e i membri di ogni coppia si ricombinano, cioè si scambiano alcune porzioni. Questo scambio, che lascia comunque ogni cromosoma completo cosicché né l'uno né l'altro dei due cromosomi perde una qualche informazione genetica, fa in modo che le singole serie di cromosomi ripartite fra le cellule germinali (ciascuna di queste cellule contiene la metà dei cromosomi delle altre cellule) rappresentino un miscuglio delle due serie parentali.

Qual è lo scopo della ricombinazione?

I cromosomi omologhi non sono affatto identici e in ciò può risiedere l'importanza della ricombinazione genetica. Ogni cromosoma in una coppia di cromosomi omologhi porta, in prevalenza, lo stesso insieme di geni, che oggi sappiamo essere sequenze di nucleotidi che codificano per molecole (di solito proteine), le quali conferiscono all'organismo forma e funzione.

Tuttavia, a causa delle mutazioni che si sono accumulate nel corso della storia evolutiva della specie, le due copie di un gene possono differire per uno o più nucleotidi e, di conseguenza, possono essere portatrici di caratteri alternativi. Una copia può addirittura essere priva di una parte o mancare del tutto, oppure la sequenza può essere interrotta da un segmento di DNA estraneo. La ricombinazione fa in modo che, in una cellula germinale, un cromosoma possa essere dotato di un qualsiasi assortimento delle mutazioni presentate dai due cromosomi parentali, e non semplicemente di una o dell'altra serie di cromosomi. In questo modo un numero illimitato di combinazioni di caratteri può venire trasmesso alla discendenza.

Dal punto di vista darwiniano la ricombinazione genetica si sarebbe evoluta proprio per questa ragione. Di fronte a un ambiente mutevole, il trasmettere combinazioni variabili di caratteri dovrebbe fornire al singolo individuo un vantaggio selettivo aumentando la probabilità che perlomeno un discendente abbia quel preciso insieme di caratteri che sono necessari per sopravvivere e per moltiplicarsi.

Evoluzionisti che sostengono un diverso punto di vista asseriscono che, nella maggior parte delle situazioni, la ricombinazione comporta uno svantaggio selettivo. Poiché il singolo individuo ha avuto successo nella riproduzione - sostengono questi studiosi - la combinazione di caratteri, specificata dai cromosomi parentali, si è, per definizione, dimostrata adattativa. A patto che l'ambiente si modifichi lentamente, quell'individuo trae maggiore beneficio «congelando» il proprio patrimonio genetico anziché rimescolandolo.

Se le cose stanno così, la ricombinazione genetica si è mantenuta solo perché deve avere qualche altra finalità. Per esempio, sembra dimostrato che la ricombinazione genetica rientri in un meccanismo di riparazione che permette a una cellula di ricostruire un cromosoma leso, prendendo in prestito l'informazione mancante da un cromosoma omologo intatto.

La descrizione che Watson e Crick hanno dato del DNA ha subito suggerito un modo per spiegare la precisione con la quale l'informazione viene scambiata tra cromosomi ricombinanti. E riflette il fatto che, nella doppia elica del DNA, le basi nucleotidiche sono disposte in paia

complementari, cosicché l'ordine con cui sono allineate le basi su un filamento della doppia elica di DNA specifica l'ordine delle basi sull'altro filamento. Se, per esempio, un segmento di DNA a filamento singolo su un cromosoma rimane scoperto nel corso della ricombinazione, può riconoscere la sequenza corrispondente su un altro cromosoma grazie alla presenza di basi complementari che si appaiano.

In che modo questo filamento di DNA su un cromosoma viene lasciato scoperto e successivamente allineato con un cromosoma omologo? Quali fasi sono necessarie a tagliare e a ricongiungere i singoli filamenti di DNA dei cromosomi che interagiscono tra loro per ricostruire uno o due cromosomi ricombinanti completi, ciascuno dei quali è una nuova combinazione delle mutazioni trasportate dai cromosomi interagenti? Questi sono gli interrogativi che mi hanno spinto, insieme ad altri ricercatori, a studiare modelli molecolari di ricombinazione più completi.

Organismi modello

A. John Clark dell'Università della California a Berkeley ha posto le basi per la costruzione dei futuri modelli molecolari quando, nel 1965, ha affrontato l'analisi genetica della ricombinazione in un batterio diffusamente studiato, l'*Escherichia coli*. Come diversi altri batteri, *E. coli* presenta una forma di riproduzione sessuale. In coppie di batteri si forma un ponte temporaneo, che permette al «maschio» di trasferire il suo unico cromosoma nella «femmina». I due cromosomi batterici hanno così la possibilità di scambiarsi alcune parti. Le ricerche di Clark e di altri hanno portato all'identificazione di geni i cui prodotti proteici catalizzano questo processo di ricombinazione.

Le mutazioni segnalavano la presenza dei geni. Nei ceppi di *E. coli* portatori di una mutazione nel primo dei tre geni da identificare, il *recA*, la ricombinazione si riduceva a un millesimo del normale. Invece, i ceppi che presentavano una mutazione nell'uno o nell'altro degli altri due geni, il *recB* e il *recC*, avevano una ricombinazione ridotta a un centesimo del normale.

Si è quindi andati in cerca delle proteine presenti nel tipo selvatico di *E. coli* (non mutante), ma assenti nei batteri mutanti. Ciò ha portato all'isolamento della proteina recA e della proteina recBC, che consiste di subunità codificate dai geni *recB* e *recC*. È stato dimostrato in seguito che questa proteina conteneva anche un terzo componente, codificato dal gene *recD*: essa è nota oggi come proteina recBCD. Non appena le proteine della ricombinazione si sono rese disponibili, si è potuta esaminare la loro interazione con le molecole di DNA, alla caccia di indizi sul loro ruolo nella ricombinazione genetica.

In provetta, la proteina recA si lega a un filamento singolo di DNA, che può essere estratto da certi piccoli virus. Il DNA così rivestito dalla proteina penetra nel DNA intatto, a duplice filamento, e fa separare i due filamenti complementari. Come una sonda, lo esplora fino a quando non trova una sequenza omologa di nucleotidi; a questo punto forma con il filamento complementare, che abbandona il partner precedente, un nuovo duplice filamento mentre la proteina recA si stacca. Ne risulta così una struttura, l'anello a D, in cui il filamento di DNA che è stato lasciato da parte, forma una curva attorno al duplice filamento (*si veda l'illustrazione nella pagina a fronte*).

Contrariamente alla proteina recA, la proteina recBCD presenta un'affinità *in vitro* per la doppia elica di DNA. In condizioni opportune, nelle microfotografie al microscopio elettronico si nota l'enzima che si insinua nella doppia elica del DNA a partire da un'estremità procedendo tra i filamenti complementari nel senso della lunghezza. Esso divide i due filamenti a mano a mano che avanza e li riunisce non appena è passato. Riassocia, tuttavia, le parti separate con più lentezza di quando le stacca, con il risultato che, da ogni parte di questo enzima in movimento, si sviluppano anelli di DNA a filamento singolo, che si ingrandiscono sempre più (*si veda l'illustrazione di questa pagina*). In alcune circostanze, la proteina recBC non soltanto separa i filamenti di DNA, ma li recide.

Per le fasi iniziali della ricombinazione

La ricombinazione di due cromosomi può essere interpretata come uno scambio tra due molecole di DNA a duplice filamento. Essa ha luogo in una regione in cui i cromosomi che interagiscono tra loro hanno una sequenza simile o identica di basi nucleotidiche, ossia delle subunità chimiche che codificano l'informazione lungo la doppia elica del DNA (*in alto*). Su ogni molecola le basi formano coppie complementari: una adenina (*A*) su un filamento si appaia sempre con una timina (*T*) sull'altro filamento, e una guanina (*G*) si appaia sempre con una citosina (*C*). Per queste regole di appaiamento, un segmento di DNA a filamento singolo proveniente da un cromosoma ricombinante può trovare la sequenza con cui appaiarsi sull'altro cromosoma. Quindi i cromosomi si allineano, con il risultato che uno scambio di segmenti tra loro porta alla formazione di due cromosomi ricombinanti completi (*in basso*). Anche se i cromosomi iniziali sono identici (o quasi) in corrispondenza del sito di ricombinazione, possono portare un'informazione genetica molto diversa in altra sede; pertanto la ricombinazione dà luogo a nuovi assortimenti di caratteri.

L'effetto della ricombinazione è schematizzato per due particelle di un batteriofago (virus che infetta i batteri) e per la loro discendenza. Le particelle virali si attaccano al batterio e vi inoculano il loro cromosoma (*a sinistra*). I due cromosomi portano varianti di geni che differiscono per il colore del rivestimento (*B* o *R*) e per la lunghezza della coda (*L* o *C*). All'interno della cellula ospite i cromosomi si replicano e si ricombinano. Nella discendenza fagica, liberata dal batterio, la ricombinazione ha riassortito i geni così da produrre tutte le possibili varianti (*a destra*).

Gli enzimi di ricombinazione del batterio *Escherichia coli* agiscono sul DNA con processi che possono promuovere la ricombinazione. In provetta, la proteina recA riveste filamenti singoli di DNA (*in alto a sinistra*). Un filamento rivestito si insinua successivamente nel DNA a duplice filamento e lo esplora fino a quando trova una sequenza nucleotidica complementare. Il filamento singolo e il filamento complementare si appaiano e formano una nuova molecola a duplice filamento (duplex), mentre si stacca la proteina recA. Il nuovo duplex e il filamento singolo che è stato appena spostato formano un «anello a D». La microfotografia al microscopio elettronico, eseguita da Charles M. Radding della Yale University, mostra uno di questi anelli, formato dalla proteina recA (*in basso a sinistra*). In vitro la proteina recBCD, cioè l'altro enzima di ricombinazione di *E. coli*, penetra nel DNA a duplice filamento a un'estremità e procede tra i due filamenti, separandoli a mano a mano che procede (*in alto a destra*). L'enzima riassocia poi i filamenti dietro di sé, ma lo fa più lentamente di quanto li separi; come risultato si formano da ogni parte dell'enzima che si sposta anelli di dimensioni crescenti di DNA a filamento singolo. Questi anelli sono visibili nella microfotografia elettronica di Andrew Taylor, che lavora nel laboratorio di Gerald R. Smith del Fred Hutchinson Cancer Research Center (*in basso a destra*). La capacità della proteina enzimatica recBCD di generare filamenti singoli di DNA e quella della proteina recA di inserirli in altri duplex dà sostegno all'ipotesi che queste proteine catalizzino gli stadi precoci della ricombinazione del cromosoma batterico.

in *E. coli*, le osservazioni biochimiche hanno consentito di formulare un'ipotesi di lavoro. La proteina recBCD separa un segmento di DNA a filamento singolo da un cromosoma. Con l'aiuto della proteina recA, il filamento libero stabilisce allora un nuovo duplice filamento con la sequenza complementare presente su un filamento di un altro cromosoma. Altre fasi, per le quali le prime osservazioni *in vitro* non hanno fornito alcun chiarimento, trasformerebbero la struttura iniziale in nuovi cromosomi ricombinati. Altri ricercatori e io siamo stati in grado di arricchire questo quadro concentrandoci non tanto su *E. coli* quanto su un virus che lo attacca, il fago lambda.

Come altri batteriofagi (virus che infettano i batteri), il fago lambda contiene un solo cromosoma, che inocula in una cellula batterica per dare inizio all'infezione. Il virus sovverte così il meccanismo biochimico del batterio per autoriprodursi, cioè per copiare il proprio DNA e per sintetizzare l'involucro proteico, o capside, all'interno del quale il nuovo cromosoma viene alloggiato dopo che il virus, moltiplicandosi, fa scoppiare la cellula ospite. Durante il breve soggiorno nell'ospite i singoli cromosomi virali si ricombinano: se due ceppi del fago, geneticamente distinguibili, sono entrambi presenti nella stessa cellula ospite, una parte della loro progenie finirà per possedere geni provenienti da entrambi i ceppi parentali.

I fagi sono soggetti interessanti per lo studio della ricombinazione grazie alla ridotta dimensione del loro unico cromosoma, alla semplice struttura delle particelle infettanti e alla rapidità con la quale svolgono le proprie attività vitali. Concentrando l'attenzione sulle infezioni provocate in *E. coli* dal fago lambda, è stato inoltre possibile esaminare gli effetti, sul singolo cromosoma di questo fago, di due sistemi di ricombinazione: quello di *E. coli* e quello dello stesso virus, che può utilizzare parti del sistema batterico, ma che è anche in grado di contribuire al processo con diversi elementi del proprio sistema.

Gli enzimi di ricombinazione del fago lambda, come quelli di *E. coli*, sono stati isolati ed esaminati per valutare la loro attività catalizzatrice *in vitro*. Uno di questi enzimi, una proteina prodotta dal gene *red-alfa*, digerisce uno dei due filamenti dei DNA a duplice filamento lasciando scoperto l'altro. La sua funzione, nella ricombinazione virale, sarebbe analoga alla funzione postulata per la proteina recBCD: fornire un DNA a filamento singolo che può andare alla ricerca del segmento omologo di un altro cromosoma. La proteina recA, prodotta dal batterio ospite, potrebbe rivestire il DNA virale a filamento singolo e dare il via alla ricerca del segmento omologo. Un secondo prodotto genico virale, la proteina red-beta, favorirebbe l'appaiamento tra il DNA a catena singola e un altro cromosoma. Gli studi *in vitro* indicano che esso accelera l'esatto appaiamento di basi tra le sequenze complementari di DNA.

La vita di un fago

L'esatta successione delle fasi di ricombinazione nei cromosomi dei fagi lambda non può essere ricostruita osservando il comportamento dei singoli enzimi di ricombinazione in provetta. Il processo deve invece essere ricostruito in un sistema vivente. Le manipolazioni sperimentali che hanno permesso ad altri ricercatori e al mio gruppo di svolgere questo tipo di ricerche sono rese possibili dal ciclo riproduttivo del fago.

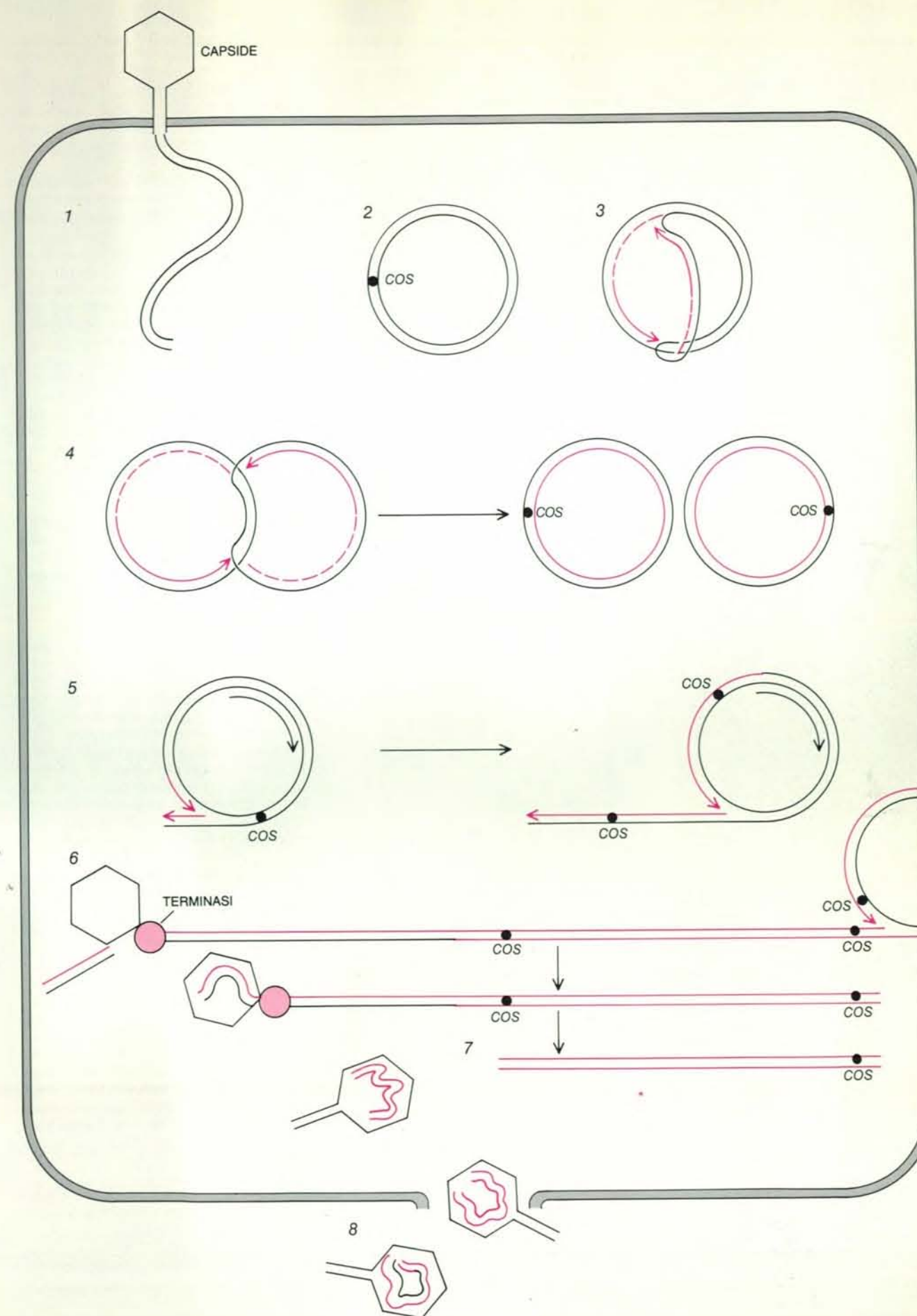
Dopo che il virus ha inoculato il suo cromosoma (un segmento di DNA contenente 48 500 coppie di nucleotidi) nella cellula di *E. coli*, le due estremità del filamento lineare si congiungono. I geni disposti lungo tutto l'anello di DNA dirigono quindi la sintesi delle proteine virali, utilizzando il meccanismo per la sintesi proteica portato dal batterio ospite. Diversi prodotti proteici sono enzimi che, operando di concerto con gli enzimi codificati dai geni dell'ospite, permettono al cromosoma del fago lambda di replicarsi. All'inizio il DNA si replica secondo una modalità che è nota come «replicazione theta». Questo tipo di replicazione produce direttamente nuovi cromosomi circolari: gli enzimi di replicazione, che operano su ciascun filamento, si fanno strada lungo il DNA parentale in direzioni opposte, generando filamenti circolari figli. Il cromosoma che si replica viene convenzionalmente rappresentato con la lettera greca theta (θ) a cui assomiglia.

Dopo diverse replicazioni theta, alcuni cromosomi cominciano a riprodursi in un modo più efficiente, noto come «replicazione sigma». L'anello di DNA si prolunga infatti in una coda, come la lettera greca minuscola sigma (σ); alla base della coda un insieme di enzimi di replicazione è continuamente attivo, riproducendo entrambi i filamenti del DNA parentale così da dare origine a una sequenza di cromosomi virali. La replicazione sigma può avere inizio quando, nella forma con replicazioni theta, compare un'intaccatura, che dà inizio alla crescita della coda. Le proteine virali, codificate da un gene chiamato *gam* e dai geni *red*, hanno un ruolo anche nella transizione alla replicazione sigma. La proteina gam inattiva la proteina ospite recBCD e una delle sue attività, che consiste nel recidere il DNA, interferisce con la replicazione secondo la modalità sigma. Non è però noto in che modo le proteine red favoriscano il passaggio da una forma all'altra.

Al momento in cui ha inizio la replicazione sigma, altri geni hanno già cominciato a dirigere la sintesi di proteine che costituiscono il capside virale e di enzimi che riempiono il capside con DNA. Il principale esponente in questa operazione di «incapsulamento» del cromosoma virale è la terminasi, un enzima che taglia il filamento cromosomico prodotto durante la replicazione sigma. In questo filamento una sequenza nucleotidica, chiamata *cos*, ricorre in quelle posizioni che corrispondono, nel cromosoma circolare, parentale, alle estremità congiunte. La terminasi riconosce la sequenza *cos* e si lega al DNA alla sua destra, se ci si riferisce al cromosoma del fago lambda come viene convenzionalmente rappresentato. La terminasi taglia quindi la sequenza *cos* alla propria sinistra.

Senza lasciar libera l'estremità tagliata, la terminasi si lega a un capside vuoto e quindi si sposta verso destra lungo il DNA, tirandosi dietro il capside e riempiendolo di DNA. Quando il complesso terminasi-capside nell'avanzare incontra una seconda sequenza *cos* non tagliata, che segna la fine di un cromosoma completo, la terminasi opera un secondo taglio e l'incapsulamento del DNA nel capside risulta completo. Se, invece, la seconda sequenza *cos* è già stata tagliata, l'incapsulamento del DNA andrà a vuoto, anche se la maggior parte del cromosoma del fago lambda è stata introdotta nel capside. Solamente i cromosomi che sono delimitati da una coppia di sequenze *cos* non tagliate possono servire per l'incapsulamento.

Dato che i cromosomi circolari, formatisi quando il DNA infettante ne ha congiunto le estremità o prodotti durante la replicazione theta, contengono soltanto un'unica sequenza *cos*, la terminasi può tagliare il DNA, ma non può in-

Il ciclo vitale del fago lambda, un virus che attacca i batteri, serve per studiare la ricombinazione. La particella virale comincia con l'attaccarsi alla membrana dell'ospite batterico, vi inocula il proprio cromosoma mentre rigetta il capside (*1*). Le estremità del cromosoma si congiungono formando un anello (*2*); in corrispondenza della giunzione si origina una sequenza nucleotidica di nome *cos*. Il DNA comincia quindi a replicarsi, dapprima in un modo non efficiente, detto theta (dalla forma della struttura che si sviluppa a un certo punto del processo), che dà origine a nuovi cromosomi circolari (*3, 4*). La replicazione si svolge poi secondo la modalità detta sigma per la rassomiglianza della struttura con la lettera greca (*5*). La replicazione sigma genera lunghi filamenti di cromosomi virali. Un enzima, la terminasi, incapsula quindi i singoli cromosomi in nuovi capsidi (che sono stati sintetizzati con l'aiuto della cellula ospite). La terminasi avvia l'incapsulamento di un cromosoma virale quando taglia il filamento cromosomico in corrispondenza di una sequenza *cos* (*6*). Si sposta quindi lungo il cromosoma, riempiendo di DNA un capside fino a quando non raggiunge una seconda *cos* (*7*). Una volta formatasi una moltitudine di nuove particelle virali, il batterio scoppia e le particelle si disperdono dando inizio a nuove infezioni (*8*).

capsularlo. Tuttavia, attraverso la ricombinazione, questi cromosomi solitari possono diventare idonei all'incapsulamento. Il principio può essere illustrato da un semplice esempio di ricombinazione tra cromosomi circolari. Nei cromosomi lineari la rottura dei filamenti di DNA e lo scambio di segmenti tra loro danno origine a due ricombinanti lineari; invece, lo stesso tipo di scambio tra due cromosomi circolari produce un'unica molecola circolare di dimensioni doppie. In questo dimero (molecola costituita da due unità), ciascuno dei nuovi cromosomi ricombinanti è ora fiancheggiato da sequenze *cos*. O l'uno o l'altro di questi cromosomi (ma non tutti e due) può ora essere incapsulato in una particella matura.

Un sistema sperimentale

La percentuale di virus ricombinanti nelle particelle liberate da un batterio che è stato infettato può essere aumentata (e l'analisi dei sistemi di ricombinazione dell'ospite e del fago facilitata) se non si lasciano replicare i cromosomi virali. Lo scopo è di impedire loro di andare incontro alla replicazione sigma, la quale genera sequenze di cromosomi che possono essere incapsulate direttamente. Dato che ogni fase del ciclo vitale del virus viene portata a termine a eccezione della replicazione del DNA, i cromosomi circolari del virus infettante possono essere di nuovo incapsulati nelle nuove particelle, ma solo se ha luogo la ricombinazione dei cromosomi.

Le mutazioni sensibili alla temperatura, che interessano i geni batterici e virali che codificano per gli enzimi che copiano il DNA, servono a bloccare la replicazione dei cromosomi virali. Queste mutazioni non si manifestano alle temperature ordinarie; altrimenti il virus e il suo batterio ospite non potrebbero propagarsi. Gli esperimenti possono essere realizzati a temperature più elevate, che attivano le mutazioni e provocano alterazioni negli enzimi della replicazione. Il fatto che i cromosomi virali non si replichino in queste condizioni può essere verificato marcando il DNA delle particelle infettanti e confrontandolo con il DNA delle particelle che fuoriescono dalla cellula ospite.

La tecnica standard per marcare il DNA del fago consiste nel coltivare il fago in un ambiente ricco di isotopi pesanti del carbonio e dell'azoto, normalmente rari. Le particelle del fago incorporano tali isotopi nel loro DNA, diventando perciò a loro volta più pesanti (più dense) dei virus coltivati in condizioni normali. Servendosi della centrifugazione in gradiente di densità, gli isotopi pesanti possono essere riconosciuti nelle successive generazioni del virus. In esperimenti in cui la replicazione sia stata bloccata, le particelle che fuoriescono dai batteri infettati con il fago marcato hanno cromosomi costituiti quasi interamente di isotopi pesanti e questo indica che, sostanzialmente, non è stato prodotto nuovo DNA.

Kenneth D. McMilin, che svolge ricerche nel mio laboratorio a proposito dei sistemi di ricombinazione del batterio ospite e del virus, ha messo a punto un metodo che combina la marcatura con isotopi pesanti con il sistema della replicazione bloccata. In tale sistema sperimentale, i batteri sono stati infettati con due tipi di particelle lambda. Un tipo era stato marcato con isotopi pesanti e portava un marcatore genetico, una mutazione facilmente identificabile, all'estremità sinistra del cromosoma. L'altro tipo conteneva isotopi leggeri, comuni, e portava una mutazione all'estremità destra. La progenie formatasi a seguito dell'infezione di un batterio da parte di particelle fagiche geneticamente e isotopicamente miste è stata poi selezionata in base alla densità.

Nella progenie virale (che dovrebbe aver subito pressoché tutta la replicazione), alcune particelle erano pesanti come le particelle infettanti marcate mentre altre erano leggere come quelle non marcate. È probabile che i due gruppi rappresentassero ricombinazioni rispettivamente tra due virus marcati e due non marcati. Le particelle la cui densità era compresa tra quegli estremi avevano presumibilmente cromosomi ricombinanti, contenenti una parte di DNA leggero e una parte di DNA pesante. I marcatori genetici su quei cromosomi, insieme con i dati sulla densità, hanno permesso di determinare dove si era concentrata la ricombinazione.

Il «sistema red»

Nell'esperimento iniziale, una mutazione nel gene *recA* di *E. coli* aveva inattivato il sistema di ricombinazione del batterio. Solamente il sistema virale, il sistema red, poteva effettuare le ricombinazioni necessarie per produrre cromosomi incapsulabili. La densità della progenie ricombinante, nella quale entrambi i marcatori erano assenti e di conseguenza includeva la parte destra di un cromosoma pesante e la parte sinistra di un cromosoma leggero, era bassa. La progenie di tipo selvatico, evidentemente, conteneva più DNA leggero che DNA pesante; pertanto la ricombinazione si era concentrata all'estremità destra del cromosoma fagico.

Quale caratteristica dell'estremità destra del cromosoma faceva concentrare in essa la ricombinazione? Le estremità del cromosoma del fago lambda sono definite dalla sequenza *cos*, la sequenza in corrispondenza della quale la terminasi può recidere singoli cromosomi circolari e i filamenti cromosomici prodotti durante la replicazione sigma. Le tecniche della clonazione di geni hanno permesso a Helios Murialdo dell'Università di Toronto e a Ichizo Kobayashi, che lavora nel mio laboratorio, di far slittare i nucleotidi che costituiscono la sequenza *cos* dalle estremità del cromosoma al centro del cromosoma stesso. Nei batteri infettati con il fago alterato, la terminasi poteva presumibilmente recidere il DNA solo in corrispondenza della nuova sede della sequenza *cos*, dimodoché, quando i cromosomi erano incapsulati, le vecchie estremità sinistra e destra diventavano il nuovo centro. Un'analisi genetica e isotopica della progenie virale, formatasi dopo l'infezione del batterio da parte del fago, ha messo in evidenza che la ricombinazione era ora concentrata in corrispondenza della nuova estremità destra del cromosoma virale, cioè in corrispondenza della nuova sede della sequenza *cos*.

Così la sequenza *cos* è un «ricombinatore», cioè una sequenza di nucleotidi che promuove nelle sue vicinanze la ricombinazione genetica. Il fatto che tale sequenza stimoli la ricombinazione solo all'estremità destra di un cromosoma (e non alla sinistra) ha fatto intravedere il suo meccanismo d'azione. Kobayashi ha proposto che la sequenza *cos* avvii un processo di ricombinazione catalizzato dal cosiddetto «sistema red» in quanto è la sede in cui la terminasi recide un cromosoma circolare. Poiché la terminasi rimane legata all'estremità sinistra del cromosoma, solamente l'estremità destra rimane a disposizione per coinvolgere un altro cromosoma. Il sistema red agisce quindi su questa estremità cromosomica libera.

Ci si aspetterebbe, dunque, che il sistema red promuova la ricombinazione dovunque i due filamenti di DNA del cromosoma del fago lambda possono essere recisi. Gli enzimi di restrizione, cioè molecole che recidono il DNA e che vengono elaborate dai batteri come arma contro i virus invasori (sono diventati anche uno strumento familiare per le manipolazioni genetiche in laboratorio), hanno permesso di verificare questa previsione. Ciascun enzima di restrizione riconosce e recide una breve sequenza nucleotidica specifica (e di conseguenza un qualsiasi cromosoma che la contenga) a meno che questa sequenza non abbia subito una modificazione chimica da parte di un altro enzima prodotto dal batterio per proteggere il proprio materiale genetico.

David Thaler, che lavora nel mio laboratorio, ha studiato la ricombinazione in batteri che producono un particolare enzima di restrizione. Un gruppo di fagi infettanti portava una sequenza bersaglio intatta, di cui era nota la posizione sul fago. Anche le altre particelle fagiche infettanti portavano la stessa sequenza, ma in una forma modificata e di conseguenza non scindibile. La modificazione assicurava che dall'ospite potessero emergere alcuni cromosomi virali completi, in grado di essere incapsulati in nuove particelle fagiche. Il sistema di ricombinazione dell'ospite veniva bloccato e rimaneva attivo solo il sistema red. Thaler ha potuto rilevare un'estesa ricombinazione, la quale appariva concentrata nella regione cromosomica che portava la sequenza amputabile. Evidentemente le estremità tagliate del cromosoma avevano cercato cromosomi non recisi e modificati e si erano ricombinati con essi.

Il ruolo della replicazione

Come si comporta il sistema red durante il ciclo riproduttivo normale del fago lambda, quando il suo DNA si replica liberamente? È difficile rispondere a questo interrogativo con la sola tecnica basata sulla separazione secondo il gradiente di densità: la replicazione senza freno del DNA porterebbe alla produzione di una tale quantità di nuovo DNA leggero che la densità delle particelle infettanti finirebbe col passare in seconda linea. Quando, al contrario, tutte le particelle infettanti contengono isotopi pesanti e la replicazione può avvenire solamente in misura ridotta, la tecnica basata sulla separazione secondo il gradiente di densità fornisce risultati del tutto inequivocabili.

I miei colleghi e io abbiamo infettato cellule batteriche con due tipi di particelle fagiche, caratterizzate da una determinata densità e contenenti svariati gruppi di marcatori genetici. Come avevamo già detto, questi marcatori hanno consentito di determinare il sito dove, nei cromosomi della progenie, aveva

L'incapsulamento di un cromosoma fagico circolare in un capside ha luogo solo se il cromosoma si ricombina. Un cromosoma circolare contiene solo una sequenza *cos*; quindi mette a disposizione un solo sito per la scissione del DNA da parte della terminasi, l'enzima che effettua un taglio iniziale e uno finale a livello di sequenze *cos* intatte per incapsulare un cromosoma. Nell'immagine, i due cromosomi ricombinanti sono circolari (*1*); l'interscambio tra loro (*2*) dà origine a una molecola circolare di dimensioni doppie (*3*), comprendente due cromosomi completi, fiancheggiati da sequenze *cos*. La terminasi inserisce in un capside uno dei cromosomi, ma non entrambi (*4*).

La distribuzione delle particelle fagiche liberate da batteri infettati in base alla densità ha fornito alcuni indizi sul sito di ricombinazione presente sul cromosoma virale. Due insiemi di particelle virali hanno dato inizio ciascuno a un'infezione. Il DNA di un insieme aveva una marcatura che lo rendeva insolitamente denso; portava, inoltre, una mutazione (*M*) all'estremità sinistra del cromosoma (*in alto a sinistra*). L'altro insieme di particelle conteneva DNA normale, con una mutazione all'estremità destra del cromosoma (*in alto a destra*). Le infezioni si effettuavano in condizioni che bloccavano la replicazione del DNA, ma lasciavano che i cromosomi ricombinanti venissero incapsulati e poi liberati. Quando il sistema di ricombinazione batterico era inattivato, e rimaneva solo il sistema di ricombinazione del virus, le particelle fagiche, i cui cromosomi non portavano né l'una né l'altra mutazione, tendevano a essere leggere (*curva in colore*). L'assenza di mutazioni indicava che il DNA era costituito dalla parte destra di un cromosoma denso e dalla parte sinistra di un cromosoma leggero; la bassa densità suggeriva che il segmento di cromosoma denso doveva essere corto. La ricombinazione deve essersi concentrata all'estremità destra del cromosoma. Quando il sistema di ricombinazione batterico era in funzione, le particelle fagiche che non portavano né l'una né l'altra mutazione esibivano un'intera gamma di densità (*curva in nero*). La ricombinazione doveva essere distribuita uniformemente lungo il cromosoma virale.

avuto luogo la ricombinazione. Quando i cromosomi consistevano interamente di DNA pesante (e pertanto non replicato), la ricombinazione era concentrata all'estremità destra, come era avvenuto negli esperimenti nei quali la replicazione era bloccata; quando, invece, consistevano in tutto o in parte di DNA leggero (sintetizzato di recente), la ricombinazione doveva aver avuto luogo in tutto il cromosoma.

Il sistema red, presumibilmente, agiva sulle estremità libere della sequenza di DNA rimaste casualmente esposte durante la replicazione. Quest'ipotesi riceve un certo appoggio da recenti studi sulla ricombinazione mediata dal sistema red in presenza di una proteina che si lega al DNA a duplice filamento e ne occupa le estremità libere: nel virus che si replica liberamente, tale proteina inibisce la ricombinazione in tutto il cromosoma. Il modo più semplice per spiegare la presenza di queste estremità libere disposte casualmente, consiste nell'attribuire la loro comparsa alla replicazione sigma. Si ritiene che la piccola intaccatura sul cromosoma circolare che dà l'avvio alla replicazione sigma compaia casualmente così come la coda che comincia a crescere in quel punto, in qualunque zona della sequenza del DNA.

L'associazione della ricombinazione mediata dal sistema red con le estremità del cromosoma, assieme al comportamento *in vitro* degli enzimi della ricombinazione red, fornisce una chiara immagine delle fasi iniziali del processo (*si veda l'illustrazione in questa pagina*). La proteina prodotta dal gene *red-alfa* digerisce in parte uno dei due filamenti di un cromosoma, a cominciare dall'estremità libera. Tale processo lascia libero DNA a filamento singolo, che invade un secondo cromosoma circolare, formando un anello a D dove trova la sequenza omologa. La proteina red-beta del fago e la proteina recA del batterio ospite possono facilitare questo evento. La separazione della struttura, formatasi per invasione di un cromosoma ricombinante da parte del DNA a filamento singolo, richiede ulteriori processi enzimatici che non sono ancora stati completamente chiariti.

Il sistema di ricombinazione del fago lambda agirebbe, secondo l'opinione più accreditata, su un'estremità libera del DNA a duplice filamento. Nel disegno la terminasi genera l'estremità libera, tagliando un cromosoma circolare del fago in corrispondenza della sequenza *cos* (*1*); l'enzima rimane poi legato all'estremità sinistra (*2*). Il prodotto proteico del gene *red-alfa* del fago digerisce in parte un filamento del cromosoma, a partire dall'estremità destra libera (*3*). L'altro filamento, che rimane solo, invade un secondo cromosoma, formando un nuovo duplex e creando un anello a D (*4*). Gli enzimi tagliano e ricongiungono i filamenti in questo anello dimodoché il cromosoma iniziale e il nuovo duplex risultano continui (*5*). La terminasi si sposta lungo il DNA, incapsulandolo in un capside vuoto (*6*). In corrispondenza della sequenza *cos* del cromosoma invaso essa lo taglia, fornendo una nuova particella virale con un cromosoma ricombinante (*7*). Il modello concilia il comportamento osservato della proteina red-alfa con la scoperta che la ricombinazione catalizzata dal sistema fagico è concentrata vicino alle estremità del duplice filamento.

Il sistema rec

I miei collaboratori e io abbiamo utilizzato gli stessi mezzi per stabilire il sito della ricombinazione mediata dal sistema red di *E. coli*. Batteri con un sistema di ricombinazione intatto sono stati infettati con fagi marcati e no, che erano portatori di mutazioni a carico dei geni *red* e *gam*, mutazioni che non solo invalidavano il sistema di ricombinazione del virus stesso, ma facevano sì che il virus non fosse più in grado di bloccare il sistema ospite. L'effetto sul cromosoma del fago è stato interessante per la diversità rispetto al risultato ottenuto con il sistema red; negli esperimenti a replicazione bloccata la ricombinazione ha avuto luogo con una frequenza pressoché uguale in tutto il cromosoma. Non si concentrava vicino alla sequenza *cos* e il modo e la frequenza con cui si svolgeva non erano diversi quando avveniva una certa replicazione del DNA. Evidentemente il sistema rec, a differenza del sistema red, non presenta una particolare affinità per le estremità recise del cromosoma.

Un altro filone di ricerca ha fatto inaspettatamente luce sul sistema rec. David Henderson dell'Università di Edimburgo stava studiando i mutanti del fago lambda che erano privi dei geni *red* e *gam*. Poiché questi geni sono essenziali non solo per la ricombinazione, ma anche per la replicazione sigma, i mutanti potevano effettuare solo la replicazione theta. Henderson notò (come avevano fatto altri ricercatori) che il virus mutante non poteva svilupparsi in batteri che avevano una mutazion e nel gene *recA*. In assenza sia del sistema di ricombinazione batterico, sia di quello virale, nessuno dei cromosomi originali o dei loro discendenti ad anello, ottenuti per replicazione theta, poteva essere incapsulato e liberato all'interno di una nuova particella.

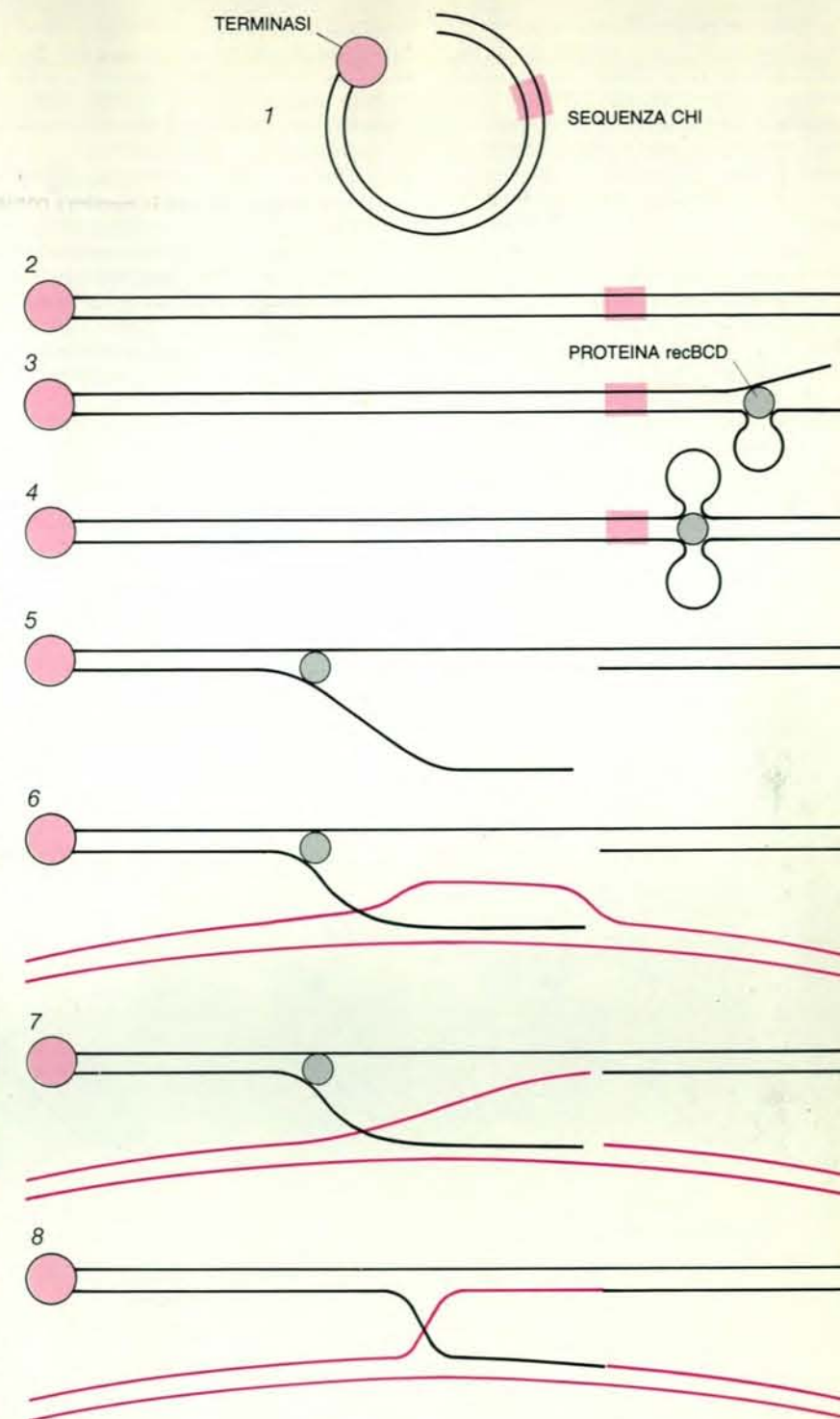
Con sua sorpresa, Henderson trovò che il mutante si sviluppava poco anche in batteri con sistema rec intatto. Venivano prodotte poche particelle virali mature anche se un sistema rec funzionante avrebbe dovuto rendere incapsulabili i cromosomi circolari mediante ricombinazione. Il sistema rec, a quanto pare, non agiva in modo efficiente sul cromosoma virale.

Un nuovo mutante secondario comparve spontaneamente nelle colture di fagi di Henderson e superò molto rapidamente il mutante originario, che sviluppava più lentamente. Il nuovo fago cresceva così bene non perché avesse qualche particolarità che lo rendeva idoneo all'incapsulamento, ma perché il sistema rec agiva sul suo DNA in modo più efficiente. La mutazione responsabile dava origine a una sequenza di otto nucleotidi, chiamata chi.

I miei collaboratori e io abbiamo utilizzato il metodo McMilin per determinare la distribuzione dei siti di ricombinazione in ceppi chi-positivi. Evidentemente abbiamo trovato che ciascuno di questi mutanti aveva uno straordinario tasso di ricombinazione vicino alla sequenza chi. Una mutazione nel cromosoma del fago lambda ci aveva fornito una sequenza che agiva come ricombinatore nel sistema rec dell'ospite batterico, proprio come fa la sequenza *cos* nel sistema red.

Può rapidamente essere chiarito in che modo la sequenza chi fa sentire il suo effetto. Qualunque modello di ricombinazione stimolata da questa sequenza deve tener conto delle sue svariate caratteristiche riconosciute dopo la sua scoperta. Per esempio si sa che la sequenza chi svolge una funzione di estrema importanza nella ricombinazione, catalizzata dal sistema rec, del cromosoma di *E. coli*, che è coperto di sequenze chi. Non ha, invece, alcun effetto su altri sistemi, come il sistema red del fago lambda. Pertanto si dovrebbe supporre che interagisca direttamente con uno degli enzimi di ricombinazione del sistema rec.

Il sistema chi sembra anche esercitare i suoi effetti a una certa distanza. Esso deve essere presente su uno soltanto dei due cromosomi interagenti per poter stimolare la ricombinazione. Sull'altro cromosoma possono anche mancare il sito corrispondente e migliaia di nucleotidi su uno dei filamenti, eppure la ricombinazione stimolata dalla sequenza chi ha luogo dove ricomincia l'omologia tra i due cromosomi. È significativo il fatto che questi effetti a distanza siano stati osservati inizialmente solo dalla parte della sequenza chi o, meglio, alla sinistra di questa sequenza (secondo la rappresentazione convenzionale del cromosoma del fago lambda).

Gli effetti della sequenza chi sono non soltanto direzionali, ma anche correlati alla sua orientazione nel cromosoma del fago lambda. Quando, ricorrendo alle tecniche dell'ingegneria genetica, la sequenza chi viene invertita all'interno del cromosoma, perde la sua funzionalità. Tuttavia, in modo sorprendente, se una sequenza *cos* invertita viene introdotta in un cromosoma lambda che porta una sequenza chi anch'essa invertita e silen-

Il sistema di ricombinazione di *E. coli* dipende, a quanto si ritiene, dall'interazione della proteina recBCD del batterio con una sequenza di DNA, di nome chi. L'illustrazione mostra un modello di interazione, così come si potrebbe svolgere nel cromosoma del fago lambda, che infetta il batterio *E. coli*. Dopo che la terminasi ha tagliato il cromosoma circolare del fago aprendolo (*1*), così da dare DNA lineare (*2*), la proteina enzimatica recBCD penetra nel DNA a duplice filamento e procede lungo di esso (*3*), continuando a formare anelli di DNA a filamento singolo fino a quando raggiunge la sequenza chi (*4*). Questo incontro finisce con il modificare il carattere dell'enzima, il quale immediatamente taglia uno degli anelli (*5*). Con il contributo della proteina recA (non raffigurata), il filamento libero invade un altro cromosoma fagico, formando un anello a D (*6*). In ulteriori fasi del processo, catalizzate da enzimi, il filamento che nell'anello a D è stato spostato viene tagliato e quindi liberato per potersi appaiare con il cromosoma iniziale (*7*). I filamenti appena giustapposti si uniscono per le estremità (*8*), formando un intermedio di Holliday, una struttura a filamenti incrociati che potranno essere separati in seguito in due cromosomi ricombinanti. Recenti risultati hanno però messo in dubbio questo modello.

te, questa sequenza diventa di nuovo attiva. La sequenza *cos* è caratterizzata dal fatto di fungere da sito per il taglio del DNA a duplice filamento da parte della terminasi. Si è osservato che quando un fago con una sequenza chi invertita silente infettava un ospite che produceva un enzima di restrizione capace di recidere il cromosoma del fago, la sequenza chi veniva di nuovo attivata. Una volta che sia attivata dal taglio del DNA a doppio filamento, una sequenza chi invertita stimola la ricombinazione alla sua destra.

Il modo migliore per spiegare queste osservazioni è attribuire alla sequenza chi la possibilità di interagire con un tipo di meccanismo mobile di ricombinazione. Questo meccanismo - costituito da uno o più enzimi - deve legarsi a un'estremità libera del cromosoma e quindi spostarsi in un unico senso lungo il DNA, fino a incontrare una sequenza chi che abbia la giusta orientazione. L'orientazione convenzionale è quella attiva quando il meccanismo di ricombinazione comincia i suoi spostamenti all'estremità libera generata di solito allorché la terminasi recide una sequenza *cos*. Una sequenza chi invertita sarà attiva soltanto se il meccanismo di ricombinazione si sposta nell'altro senso, cioè dalla estremità sinistra del cromosoma. La terminasi genera un'estremità sinistra libera solo se la sequenza *cos*, che segna il punto in cui l'enzima taglia un cromosoma circolare, è anch'essa invertita.

Una volta che il meccanismo mobile di ricombinazione abbia incontrato una sequenza chi opportunamente orientata, le sue probabilità di agire aumentano. In conformità con l'osservazione che la sequenza chi è in grado di agire a distanza, l'evento di ricombinazione può svolgersi dopo che è stata superata la sequenza chi: alla sua sinistra, se la sequenza è orientata in modo convenzionale; alla sua destra, se la sequenza risulta invertita.

L'intermedio di Holliday

Il meccanismo mobile di ricombinazione del sistema rec non solo è stato descritto, ma si sono fatte alcune ipotesi sulla sua identificazione. Ha proprietà che ricordano il comportamento *in vitro* dell'enzima recBCD: un enzima che impegna un'estremità libera del DNA a duplice filamento, si sposta lungo di esso e genera, durante il percorso, filamenti singoli di DNA. Inoltre, sono state identificate mutazioni del gene responsabile di recBCD, che lasciano il sistema rec attivo, ma insensibile alla sequenza chi. Esse suggeriscono un'interazione diretta tra la proteina codificata dal gene *recBCD* e la sequenza chi.

Come può la proteina agire da meccanismo di ricombinazione e come può la sequenza chi stimolarla? Queste domande possono divenire meno teoriche se si prende in considerazione una struttura intermedia che si generi quando la ricombinazione ha luogo lontano dall'estremità di un DNA a duplice filamento, come avviene nella ricombinazione catalizzata dal sistema rec. Nel 1964, Robin Holliday, allora al John Innes Institute di Hertford, in Inghilterra, ha proposto che, in qualche fase della ricombinazione, il processo di scambio di segmenti sul DNA avvenga a carico di un solo filamento in ogni cromosoma. Il filamento finisce così con l'avere le estremità recise e quindi ricongiunte e la sua prosecuzione sull'altro cromosoma dà origine a una «struttura a filamenti incrociati», chiamata anche intermedio di Holliday.

La separazione di questo intermedio di Holliday in cromosomi distinti richiede un altro evento di scissione e ricongiungimento dei filamenti cromosomici. A causa della geometria della struttura intermedia (*si veda l'illustrazione in questa pagina*), questo evento può svolgersi o sui filamenti che avevano subito originariamente scambi o sui loro filamenti complementari. Nel primo caso la separazione conduce a «ricombinazione con rattoppo», cioè cromosomi ricombinanti essenzialmente immutati, che hanno scambiato semplicemente un breve segmento (il rattoppo) di DNA a filamento singolo. Nel secondo caso le catene complementari sono andate anch'esse incontro a uno scambio di segmenti, completando così il quadro degli scambi tra cromosomi e formando «ricombinanti *cross-over*».

La domanda si pone allora in questi termini: la proteina recBCD, attivata da chi, catalizza la formazione degli intermedi di Holliday oppure la loro separazione? Un'osservazione, effettuata da Gerald R. Smith e collaboratori al Fred Hutchinson Cancer Center a Seattle, sembra convalidare il modello basato sulla prima possibilità. Il gruppo di Seattle ha trovato che in certe condizioni la proteina, spostandosi *in vitro* lungo una molecola di DNA, reagisce a una sequenza chi incontrata con l'orientazione giusta, scindendo un filamento particolare: quello che, durante la sintesi delle proteine virali, viene trascritto, o letto, da destra a sinistra.

Il modello ammette che la scissione interrompa la riassociazione del DNA, che ha generalmente luogo alle spalle dell'enzima che si sposta. Il filamento singolo che ora ha una estremità libera viene ricavato dal DNA a duplice filamento. Esso si riveste della proteina recA e invade il cromosoma omologo, formando un anello a D. Il filamento che, nella formazione di questo anello, viene spostato incrocia il cromosoma originario e una successione di tagli e di ricongiunzioni portano a un intermedio di Holliday (*si veda l'illustrazione nella pagina a fronte*). La sua separazione può dare origine o a ricombinanti con rattoppo o a ricombinanti cross-over. Questo modello che implica l'azione della sequenza chi e della proteina enzimatica recBCD, è così verosimile che la più recente edizione di un libro di testo mol to diffuso lo presenta come una realtà. Tuttavia, un recente lavoro sembra invalidarlo.

Susan Rosenberg, che lavora nel mio laboratorio, ha voluto sottoporre a verifica una specifica previsione del modello. Quest'ultimo ammette che lo stesso filamento cromosomico scisso *in vitro* dalla proteina recBCD (si tratta del filamento che viene letto da destra a sinistra) venga scisso anche *in vivo* e serva per avviare la ricombinazione. Se questo filamento letto da destra a sinistra e scisso invade un cromosoma omologo e si appaia con una sequenza complementare di DNA, finirà per spostare quel filamento del cromosoma che viene letto da destra a sinistra. Pertanto, nell'intermedio di Holliday entrambi i filamenti incrociati saranno filamenti letti da destra a sinistra. Il modello prevede quindi che, se l'intermedio si risolve in ricombinanti a rattoppo, questi ultimi compariranno sui filamenti cromosomici che vengono letti da destra a sinistra.

Susan Rosenberg, invece, ha trovato esattamente l'opposto. Ogni rattoppo di DNA compariva su un filamento letto da sinistra a destra, cioè su quel filamento che non viene scisso *in vitro*. Questo risultato mette quindi fuori giuoco il semplice modello basato sull'osservazione *in vitro*. Anche un'altra versione del modello ammette che un filamento letto da destra a sinistra e reciso avvii la ricombinazione, ma descrive uno scenario più complesso in cui i rattoppi vanno a finire sul filamento del cromosoma del fago lambda che viene letto da sinistra verso destra. Questo secondo modello prevede determinati comportamenti, uno dei quali è stato controllato dalla Rosenberg e trovato falso.

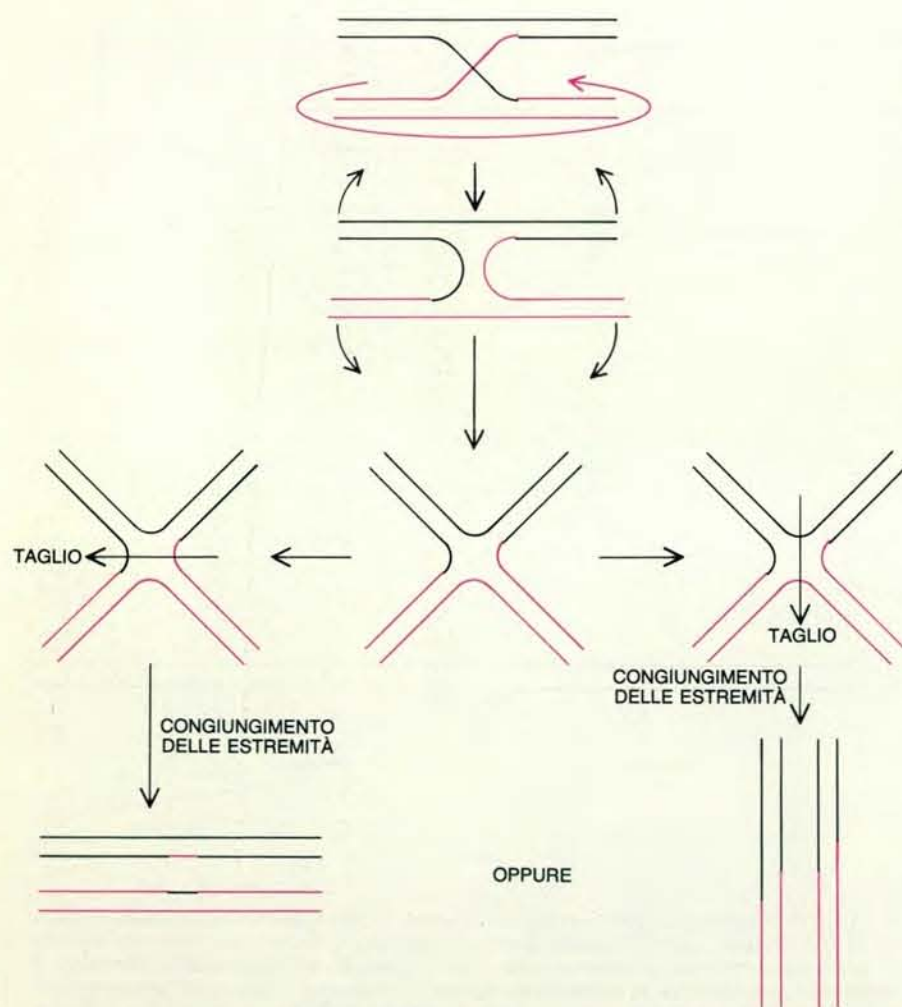
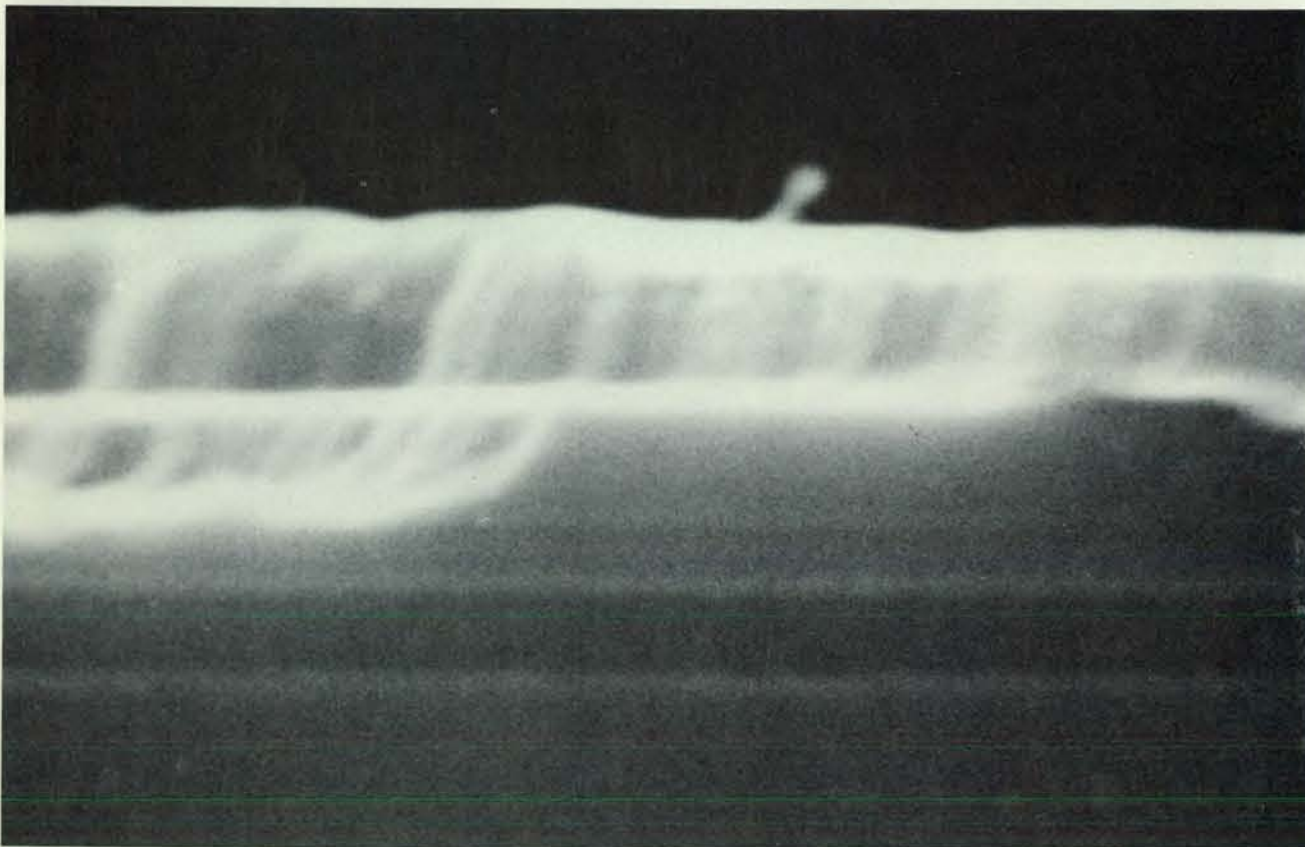
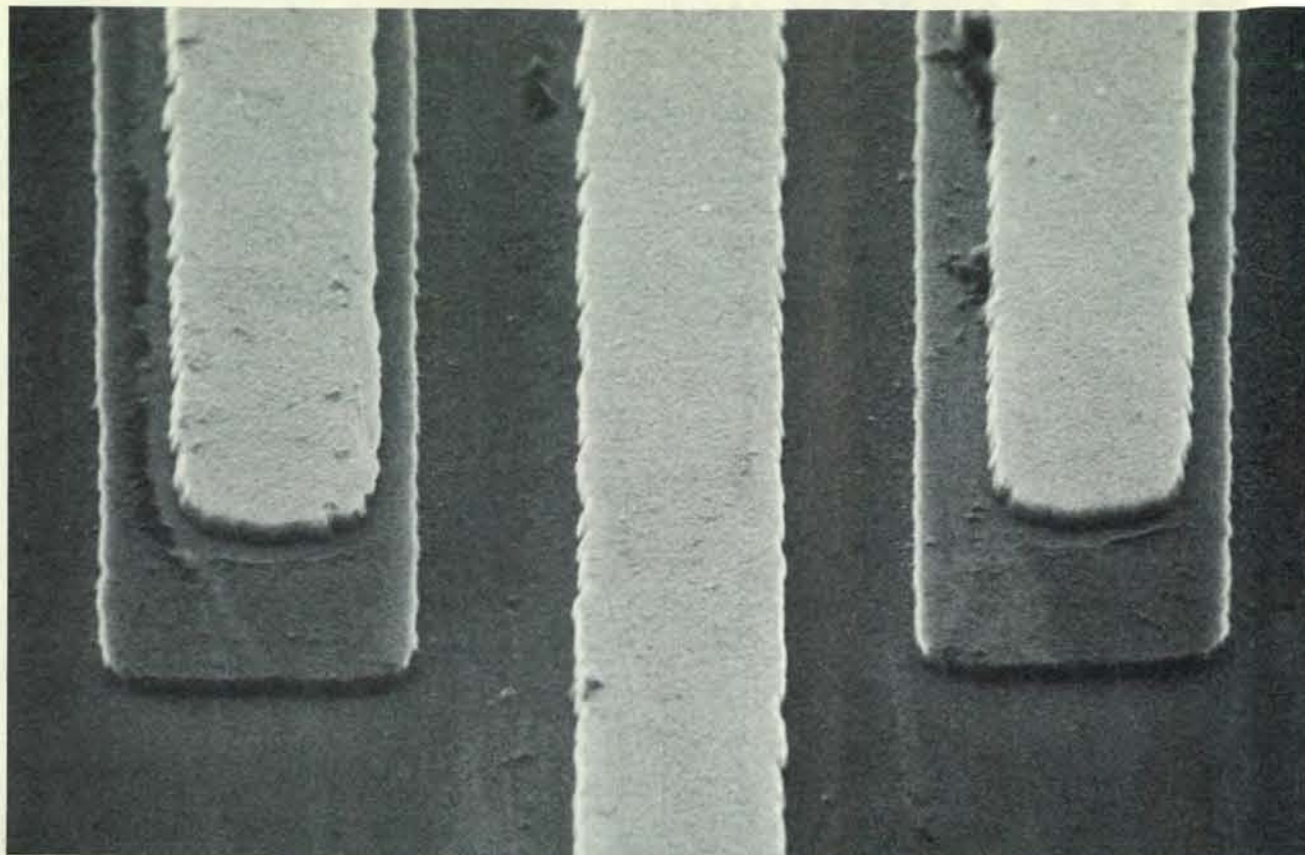
L'inadeguatezza dei due modelli suggerisce che la scissione del filamento, stimolata *in vitro* dalla sequenza chi, pur avendo luogo anche *in vivo*, potrebbe non servire ad avviare la ricombinazione. Si ritorna così a chiederci se la sequenza chi stimoli davvero la formazione o la separazione degli intermedi di Holliday. Gli sforzi fatti per rispondere a questa domanda e per caratterizzare l'interazione tra la sequenza chi e la proteina recBCD, che ne è alla base, stanno stimolando nuove ricerche.

Varianti della ricombinazione

I particolari rimangono ancora da risolvere, ma il quadro complessivo è ormai chiaro: la ricombinazione nel fago lambda e in *E. coli* segue due vie molecolari molto diverse.

Nel sistema red del virus, l'estremità libera della molecola a doppio filamento è il sito attivo; è proprio qui che gli enzimi della ricombinazione cominciano a operare per promuovere la ricombinazione. Il sistema si adatta allo stile di vita del fago lambda, in quanto le estremità libere compaiono a caso nella sequenza cromosomica durante la replicazione sigma. Invece, nel batterio le estremità cromosomiche sono fisse; il batterio richiede, dunque, un differente sistema per permettere che la ricombinazione abbia luogo nel suo corredo genetico. Nel sistema rec del batterio un'estremità cromosomica libera serve esclusivamente come sito di carico di un meccanismo di ricombinazione, che catalizza la ricombinazione a una distanza arbitraria dalla fine.

Sembra, così, che la ricombinazione vari nei suoi meccanismi esatti secondo le peculiarità genetiche dell'organismo. Un singolo organismo può impiegare varie vie: per esempio, *E. coli* possiede un sistema ausiliario, noto come recF, e sembra anche che certi funghi ricombinino il loro materiale genetico in molti modi. Studi effettuati sulle basi molecolari della ricombinazione in organismi più complessi potranno rivelare ulteriori varianti.

Questi studi ricevono un impulso dalla prospettiva che i meccanismi naturali di replicazione potrebbero un giorno essere sfruttati per sostituire geni in organismi pluricellulari, una possibilità che potrebbe risultare vantaggiosa in agricoltura e nella terapia delle malattie genetiche. L'inserzione di un gene alterato nel punto opportuno su un cromosoma esige una precisione che va ben oltre quella delle attuali tecniche di ingegneria genetica. I meccanismi raffinati e precisi che ogni organismo utilizza per ricombinare il proprio corredo genetico potrebbero benissimo diventare gli strumenti di una futura biotecnologia.

Secondo il modello proposto da Robin Holliday nel 1964, la separazione di una struttura a filamenti incrociati può dare risultati contrastanti. La struttura, formatasi precocemente in alcuni processi di ricombinazione (*si veda l'illustrazione nella pagina precedente*) può essere rappresentata con una configurazione «a crocevia», in cui quattro segmenti di DNA a duplice filamento si intersecano ad angoli retti. Il taglio a livello dell'intersezione e il congiungimento delle estremità a filamento singolo da ogni parte del taglio possono produrre «ricombinanti a rattoppo», cioè cromosomi che si sono scambiati non più di un corto segmento di DNA a filamento singolo (*a sinistra*). Una alternativa del processo, sempre con taglio e congiungimento di filamenti cromosomici, può produrre «ricombinanti cross-over», cioè cromosomi ricombinanti maturi (*a destra*).

Un dispositivo a elettroni balistici, nel quale gli elettroni possono muoversi attraverso un sottile canale senza essere diffusi dagli urti, visto dall'alto (*in alto*) e di fianco (*in basso*). Il dispositivo è formato da un certo numero di sottili strati di semiconduttore che ne costituiscono gli elementi, tra i quali l'emettitore di elettroni (vicino alla sommità del dispositivo), il canale attraverso il quale gli elettroni passano e la regione nella quale vengono raccolti. Il dispositivo è ingrandito circa 10 000 volte nell'immagine superiore e circa 100 000 volte in quella inferiore. Le fotografie sono state fatte da Michael Rosenfield, Carmello Aliotta e Gordon Wilson presso il Thomas J. Watson Research Center della IBM.

Gli elettroni balistici nei semiconduttori

Si muovono senza subire diffusione in dispositivi che già consentono di approfondire le conoscenze delle leggi della meccanica quantistica e che potrebbero in futuro rivelarsi molto più veloci dei componenti odierni

di Mordehai Heiblum e Lester F. Eastman

Nel cinescopio di un televisore gli elettroni si trasferiscono dal catodo allo schermo con moto balistico: seguono cioè traiettorie regolari che non sono interrotte da interazioni o collisioni con altre particelle. Dato che nel cinescopio le molecole di gas sono pochissime, esse diffondono pochi elettroni. Nei dispositivi a semiconduttore come i transistori, viceversa, il moto degli elettroni che trasportano la corrente è continuamente interrotto dalla diffusione. Ciascun elettrone è diffuso più volte e in ogni direzione, tanto che il suo moto somiglia a quello casuale di una molecola in un gas o in un liquido ordinari. In media un elettrone percorre con questo moto casuale un cammino molto più lungo della distanza lineare che supera. Il flusso risultante di elettroni nella direzione del campo elettrico è pertanto molto più lento della velocità effettiva degli elettroni. Se fosse in qualche modo possibile far muovere gli elettroni nei semiconduttori senza fenomeni di diffusione, cioè con moto balistico, allora la velocità dei transistori potrebbe in teoria aumentare drasticamente.

Ebbene, tale moto balistico è stato ottenuto recentemente. Le nostre ricerche presso il Thomas J. Watson Research Center della IBM, a Yorktown Heights, e il lavoro di ricercatori di altri istituti hanno permesso di determinare le condizioni nelle quali gli elettroni possono avere moto balistico e hanno portato di recente alle prime rilevazioni inequivocabili del moto balistico nei dispositivi a semiconduttore.

Il nostro interesse per gli elettroni balistici è di natura sia tecnica, sia scientifica. Sotto il profilo tecnico, gli elettroni balistici potrebbero costituire il cardine di una generazione di dispositivi elettronici più veloci dei dispositivi odierni di almeno un ordine di grandezza. Sotto il profilo scientifico, i dispositivi a elettroni balistici consentono di approfondire lo studio delle proprietà quantomeccaniche degli elettroni coerenti nei semiconduttori, perché gli elettroni balistici nei semiconduttori, come quelli nel cinescopio, possono essere studiati come pacchetti d'onda quantomeccanici in moto libero; essi manifestano effetti ondulatori (come l'interferenza) che di solito si osservano nelle onde elettromagnetiche.

In una valvola termoionica (costituita, come noto, da un tubo a vuoto), gli elettroni possono manifestare il moto balistico perché la distanza che devono superare all'interno della valvola è più piccola del loro cammino libero medio (che è la distanza che un elettrone percorre in media tra due successive collisioni con le rare molecole di gas contenute nella valvola). Nei semiconduttori è impossibile eliminare la diffusione in modo così semplice. Per quanto sorprendente, la densità del mezzo solido non costituisce di per sé un impedimento al moto balistico, benché gli atomi di un semiconduttore siano diecimila volte più vicini l'uno all'altro che le molecole di gas in una valvola termoionica. Gli atomi del semiconduttore, benché siano così densamente impaccati, non diffondono necessariamente gli elettroni, perché gli atomi di un solido cristallino sono disposti in un reticolo periodico: si trovano a distanze uguali secondo una configurazione che si ripete. La meccanica quantistica ci insegna che, per certi aspetti, un elettrone può essere trattato, sotto il profilo matematico, come un'onda; grazie a questo comportamento di tipo ondulatorio, un elettrone non viene diffuso quando interagisce con una configurazione cristallina perfettamente periodica di atomi. Questo risultato sorprendente, dimostrato da Felix Bloch nel 1928, costituisce uno dei capisaldi della fisica dello stato solido. (L'elettrone, tuttavia, non si comporta come un elettrone libero nel vuoto, bensì come una particella più leggera, con una «massa efficace» minore di quella dell'elettrone libero e con una certa velocità massima; la massa efficace e la velocità massima sono imposte dal mezzo cristallino nel quale l'elettrone si muove.)

In un cristallo reale a qualunque temperatura superiore allo zero assoluto, l'ordine periodico perfetto è solamente approssimato dalla configurazione effettiva degli atomi. Per prima cosa, gli atomi del solido sono sempre soggetti all'agitazione termica (il moto vibratorio casuale dovuto all'energia termica fornita dall'ambiente). Le vibrazioni termiche generano onde di compressione e di espansione che si propagano nel cristallo e possono diffondere gli elettroni, dato che le onde disturbano la disposizione periodica perfetta degli atomi. Quando viene diffuso da una vibrazione reticolare, l'elettrone cede o acquista dal cristallo quantità fisse di energia, dette quanti. Un evento di diffusione di questo tipo di chiama urto anelastico.

Nei semiconduttori, gli elettroni possono essere diffusi non solo dalle vibrazioni reticolari, ma anche dagli ioni delle impurezze che sono sempre presenti nei materiali reali. Questi ioni hanno una distribuzione casuale e non una configurazione periodica, perciò diffondono le onde elettroniche. Gli urti con gli ioni non alterano di molto l'energia degli elettroni (perché gli ioni hanno una massa molto maggiore), ma modificano la direzione del loro moto. Tali collisioni si chiamano urti elastici. Un elettrone viene diffuso in molti altri modi, per esempio dagli urti anelastici con altri elettroni e dai difetti e dalle irregolarità nella struttura del reticolo cristallino. Tutti questi urti accorciano il cammino libero medio degli elettroni nei semiconduttori.

Nei dispositivi a semiconduttore oggi esistenti la corrente elettronica percorre distanze molte volte più grandi del cammino libero medio degli elettroni. Per consentire agli elettroni di muoversi con moto balistico, si deve o aumentare il cammino libero medio (scegliendo in modo opportuno i materiali semiconduttori e lavorandoli in modo tale che risultino estremamente puri) oppure rimpicciolire di parecchio la regione attraversata dalla corrente (ricorrendo a tecniche di fabbricazione molto raffinate).

Il complesso comportamento degli elettroni che si spostano lungo brevi percorsi può essere illustrato da un tipico dispositivo a semiconduttore: il transistore a effetto di campo. In un transistore a effetto di campo l'applicazione di una differenza di potenziale fa muovere gli elettroni da una regione chiamata sorgente a una regione chiamata pozzo, attraverso una sottile regione chiamata canale (*si veda la figura nella pagina a fronte*). Vicino al canale vi è un contatto metallico chiamato porta. Se alla porta viene applicata una tensione negativa, si genera un campo elettrico che respinge gli elettroni da una porzione del canale; in sostanza si crea un canale conduttore più stretto e con resistenza superiore e di conseguenza la corrente trova più difficoltà a passare dalla sorgente al pozzo. Il transistore può quindi fungere da amplificatore: una piccola variazione della tensione della porta può causare una cospicua variazione della quantità di corrente che attraversa il dispositivo.

Per quanto riguarda la velocità, la dimensione più importante di un transistore a effetto di campo è la lunghezza del metallo della porta. La velocità con cui gli elettroni attraversano il canale situato sotto la porta e fra la sorgente e il pozzo determina il tempo necessario perché il dispositivo risponda alle variazioni di tensione della porta. In un normale transistore a effetto di campo il canale è lungo circa un micrometro, che è da 10 a 100 volte il cammino libero medio degli elettroni (a seconda del materiale semiconduttore di cui è fatto il dispositivo): quindi il moto degli elettroni non può essere balistico.

Che cosa accade esattamente agli elettroni quando vanno dalla sorgente al pozzo con moto non balistico? Appena abbandonano la sorgente, vengono accelerati dalla differenza di potenziale tra sorgente e pozzo. A una distanza caratteristica dalla sorgente (una distanza grosso modo pari al cammino libero medio) la velocità media degli elettroni raggiunge un massimo. A distanze superiori dalla sorgente la velocità media scende lentamente verso un livello stabile. Questo fenomeno si chiama sovraelongazione della velocità. La diminuzione della velocità è provocata soprattutto dagli urti anelastici con le vibrazioni reticolari.

Se gli elettroni potessero spostarsi nel cristallo con moto balistico, all'inizio subirebbero un'accelerazione più o meno simile a quella di un elettrone sotto l'effetto di un campo elettrico nello spazio libero (benché nel semiconduttore gli elettroni abbiano una massa efficace più piccola): la loro velocità sarebbe circa proporzionale alla radice quadrata della distanza percorsa. Ma anche in assenza di urti gli elettroni non accelererebbero indefinitamente. Per la natura dell'interazione fra gli elettroni e il reticolo periodico, prima o poi essi raggiungerebbero una velocità massima. Questa velocità massima è determinata dalla struttura del cristallo, dalla composizione atomica del semiconduttore e dalla direzione del moto dell'elettrone rispetto al reticolo cristallino.

La velocità media dell'elettrone nel suo moto dalla sorgente al pozzo è compresa fra la velocità con cui esso lascia la sorgente e la velocità massima che può raggiungere nel semiconduttore. È evidente che l'elettrone avrebbe una velocità molto più elevata (e il dispositivo un funzionamento molto più rapido) se l'elettrone potesse essere «lanciato» nel canale conduttore alla massima velocità possibile. Nei dispositivi balistici che abbiamo costruito siamo riusciti a ottenere proprio un effetto del genere. Per comprendere le modalità del lancio e la costruzione dei dispositivi balistici stessi è prima necessario conoscere alcune proprietà dei materiali semiconduttori.

Il semiconduttore più comune è il silicio. In un cristallo di silicio, ogni atomo ha quattro elettroni di valenza, che appartengono al guscio più esterno (sono quelli che interagiscono nelle reazioni chimiche) e che vengono messi in comune con quattro atomi limitrofi. Gli elettroni si comportano come una colla, dando coesione al cristallo. In certi semiconduttori, detti composti, come l'arseniuro di gallio, un elemento (il gallio) mette in gioco solo tre elettroni di valenza mentre l'altro (l'arsenico) ne mette in gioco cinque, sicché il numero medio degli elettroni di legame per atomo è ancora quattro. Quando sono strettamente localizzati fra due nuclei gli elettroni non possono trasportare corrente; per potersi muovere liberamente nel cristallo, un elettrone deve ricevere una certa quantità di energia che lo trasforma da elettrone di legame (o di valenza) in elettrone di conduzione. Gli elettroni di conduzione hanno pertanto un'energia potenziale elettrica superiore agli elettroni di legame, così come un satellite in volo ad alta quota ha un'energia potenziale gravitazionale maggiore di un satellite che stia per essere lanciato. La differenza tra l'energia potenziale minima degli elettroni di conduzione in un semiconduttore e l'energia potenziale massima degli elettroni di legame si chiama banda proibita o intervallo fra le bande e la sua entità varia da un semiconduttore all'altro.

Nella maggior parte dei semiconduttori d'impiego comune, a temperatura ambiente, solo pochi elettroni hanno energia termica sufficiente per diventare elettroni di conduzione (dovrebbero avere un'energia termica maggiore dell'energia della banda proibita), e quindi la maggior parte dei semiconduttori puri conducono male la corrente. Per aumentare la conducibilità di un semiconduttore, lo si può «drogare» aggiungendovi una quantità piuttosto piccola di un altro elemento (impurezza). Per esempio è possibile drogare il silicio aggiungendovi atomi di un elemento che abbia cinque elettroni di valenza, come il fosforo o l'arsenico. Poiché per formare il legame cristallino bastano quattro elettroni per atomo, con una piccola quantità di energia il quinto elettrone può essere liberato e condurre elettricità.

La banda proibita dell'arseniuro di gallio puro è più ampia di quella del silicio puro, quindi a temperatura ambiente il primo è un conduttore ancora peggiore del secondo. È tuttavia possibile drogare l'arseniuro di gallio con certi elementi, come il silicio, per renderlo più conduttore; gli atomi di silicio sostituiscono alcuni atomi di gallio; uno dei quattro elettroni del silicio viene liberato e contribuisce alla conducibilità elettrica. Fra tutti i semiconduttori, l'arseniuro di gallio è uno dei più idonei alla costruzione di dispositivi a elettroni balistici, come uno di noi (Eastman) ha messo in evidenza per la prima volta nel 1977. Nell'arseniuro di gallio, gli elettroni hanno una velocità massima maggiore che nel silicio (nell'arseniuro di gallio la velocità massima è circa 10^8 centimetri al secondo, mentre nel silicio è circa 2×10^7 centimetri al secondo) ed è minore la probabilità che urtino con le vibrazioni reticolari. Nell'arseniuro di gallio moderatamente drogato il cammino libero medio di un elettrone è di circa 100 nanometri (un nanometro è pari a 10^{-9} metri), grosso modo dieci volte il cammino libero medio degli elettroni nel silicio.

Un semiconduttore che viene spesso

La diffusione degli elettroni nei semiconduttori è dovuta alle vibrazioni termiche (causate dal calore) del reticolo cristallino del semiconduttore e agli atomi delle impurezze presenti in qualsiasi campione reale di semiconduttore. Gli atomi di un cristallo ideale (*a*) sono disposti secondo una configurazione perfettamente periodica. Avendo ceduto elettroni per assicurare la coesione del cristallo, possiedono una carica positiva, perciò generano un potenziale elettrico dotato di periodicità spaziale.

La meccanica quantistica insegna che gli elettroni possiedono proprietà ondulatorie. Le onde elettroniche interagiscono con il potenziale periodico in maniera tale da attraversare il cristallo ideale senza subire alcuna diffusione. In un cristallo reale le vibrazioni termiche provocano onde di compressione e di espansione che disturbano il potenziale periodico (*b*). Queste regioni di potenziale irregolare in movimento e gli atomi delle impurezze immersi nel cristallo diffondono gli elettroni.

I transistori a effetto di campo possono essere costruiti secondo varie configurazioni. In una configurazione tipo (*in alto*), gli elettroni si muovono, attraverso un canale conduttore, da una regione chiamata sorgente a un'altra chiamata pozzo, entrambe fatte di semiconduttore di elevata conducibilità. Quando a una «porta» metallica sovrastante il canale viene applicata una tensione negativa, gli elettroni sono respinti da una regione del canale chiamata strato di svuotamento. Il canale conduttore diviene quindi più stretto e la corrente diminuisce. Il transistore è un amplificatore: una piccola variazione della tensione applicata alla porta provoca una forte variazione della corrente di elettroni. La rapidità con cui il transistore reagisce alle variazioni della tensione è in rapporto diretto con la velocità alla quale gli elettroni possono spostarsi dalla sorgente al pozzo. Nel transistore a effetto di campo verticale (*in basso*), che è una configurazione recente, i vari elementi sono costituiti da sottili pellicole fatte crescere una sopra l'altra. E' pertanto possibile rendere molto breve il tragitto degli elettroni dalla sorgente al pozzo (e molto rapida la reazione del dispositivo) semplicemente realizzando il canale con una pellicola estremamente sottile. La facilità con cui si può costruire un canale conduttore sottilissimo rende il transistore a effetto di campo verticale adatto all'impiego nei dispositivi a elettroni balistici.

impiegato insieme con l'arseniuro di gallio è la lega di arseniuro di gallio e alluminio. Quando gli atomi di alluminio sostituiscono quelli di gallio nel reticolo cristallino, non viene liberato alcun elettrone: l'alluminio e il gallio hanno infatti lo stesso numero di elettroni di valenza. Gli atomi di gallio possono essere sostituiti da quelli di alluminio in proporzione arbitraria. Poiché la struttura reticolare dell'arseniuro di gallio e alluminio è simile a quella dell'arseniuro di gallio, è possibile far crescere un cristallo di uno di questi semiconduttori direttamente sulla superficie dell'altro in modo che i cristalli si attaglino in modo pressoché perfetto. Una struttura di questo genere si chiama eterogiunzione.

Una delle differenze più importanti fra l'arseniuro di gallio e l'arseniuro di gallio e alluminio riguarda l'ampiezza della loro banda proibita. Nelle leghe arseniuro di gallio e alluminio, la larghezza della banda proibita è direttamente proporzionale alla quantità di alluminio che contengono, sicché le leghe che contengono una percentuale elevata di alluminio hanno bande proibite più ampie dell'arseniuro di gallio puro o delle leghe di arseniuro di gallio e alluminio che contengono poco alluminio. Quindi, in un'eterogiunzione, gli elettroni di conduzione situati dalla parte dell'arseniuro di gallio e alluminio hanno un'energia potenziale maggiore degli elettroni che si trovano dalla parte dell'arseniuro di gallio. Quando gli elettroni di conduzione passano dall'arseniuro di gallio e alluminio all'arseniuro di gallio, il loro eccesso di energia potenziale si trasforma di colpo in energia cinetica ed essi vengono accelerati quasi istantaneamente a velocità elevatissime.

Un'eterogiunzione di questo tipo può fungere da iniettore di elettroni balistici. Essa rappresenta una delle soluzioni al problema di imprimere agli elettroni una velocità elevata prima che essi entrino nel canale di un transistore, dove dovrebbero muoversi con moto balistico. Altri due tipi di iniettori possono lanciare gli elettroni con energie superiori o regolabili. Si tratta della cosiddetta barriera a piano drogato (una delle cui varianti si chiama barriera a gobba di cammello), e della barriera a effetto tunnel o semplicemente barriera a tunnel.

In una barriera a piano drogato un piano di ioni di carica negativa (per esempio atomi di berillio - che normalmente hanno solo due elettroni di valenza - con un elettrone supplementare) viene inserito in un sottile campione di arseniuro di gallio non drogato, che a sua volta viene introdotto fra due regioni di arseniuro di gallio fortemente drogato con un'elevata densità di elettroni (*si veda l'illustrazione a pagina 56*). Gli atomi del piano drogato (il piano del berillio) formano una barriera di cariche elettriche negative che tendono a respingere gli elettroni di conduzione. Questi devono quindi possedere un'energia minima per poter superare la barriera e passare dall'una all'altra delle regioni di arseniuro di gallio drogato.

Sotto il profilo energetico, il piano di berillio rappresenta un picco di energia potenziale. L'energia potenziale parte da un valore basso nella regione fortemente drogata a un capo del dispositivo, aumenta nella regione non drogata di arseniuro di gallio, raggiunge un picco in corrispondenza del piano di ioni negativi e decresce verso l'altro capo della regione non drogata, terminando con un valore basso nella seconda regione fortemente drogata. L'altezza del picco (cioè la quantità di energia necessaria per superarlo) è proporzionale al numero degli ioni negativi e può anche uguagliare quella della banda proibita del materiale. (In una barriera a gobba di cammello, il berillio è disperso in una stretta regione, anziché essere confinato in un unico piano, e quindi il picco di energia si addolcisce in una gobba). Collocando il piano del berillio più vicino a un'estremità che all'altra dello strato di arseniuro di gallio non drogato, si può creare una barriera molto asimmetrica, che ha un fianco lungo e poco inclinato e l'altro breve e ripido. Se al fianco più lungo della barriera si applica una tensione negativa, gli elettroni collocati da questa parte acquistano un'energia potenziale maggiore e si spostano verso il piano di ioni negativi. È come se il pendio più dolce della barriera fosse stato sollevato e reso ancora più dolce. Alcuni elettroni acquistano abbastanza energia termica da attraversare il piano di ioni negativi. Scendendo il fianco ripido della barriera di energia potenziale, gli elettroni vengono accelerati rapidamente e iniettati ad alta velocità nella vicina regione di arseniuro di gallio drogata, purché il fianco ripido sia molto più corto del cammino libero medio degli elettroni.

I diagrammi dei livelli energetici dei semiconduttori permettono di spiegarne molte proprietà. Gli elettroni di valenza, messi in comune dagli atomi per assicurare la coesione del cristallo, sono localizzati fra gli atomi stessi e non possono muoversi per trasportare corrente. Per diventare elettroni di conduzione devono ricevere una certa quantità di energia. La differenza tra l'energia degli elettroni di conduzione e quella degli elettroni di valenza, la banda proibita, varia da un semiconduttore all'altro. La banda proibita dell'arseniuro di gallio (*a*), per esempio, è più stretta di quella dell'arseniuro di gallio e alluminio (*b*). La conducibilità di un semiconduttore può essere accresciuta «drogandolo», cioè aggiungendogli atomi che possano cedere facilmente elettroni, avendone più di quelli necessari al legame (*c*). Gli elettroni degli atomi dell'agente drogante (*in colore*) richiedono meno energia per diventare elettroni di conduzione lasciando gli atomi di impurezze con carica positiva.

Per costruire l'altro tipo di iniettore di elettroni balistici, la barriera a tunnel, si inserisce un sottile strato di arseniuro di gallio e alluminio non drogato fra due regioni di arseniuro di gallio drogato contenente un'elevata densità di elettroni. Normalmente l'arseniuro di gallio e alluminio agisce da barriera e gli elettroni non possono attraversarlo. Se tuttavia lo spessore della barriera è inferiore ai 10 nanometri circa, vi è una certa probabilità che gli elettroni l'attraversino nei due sensi grazie al fenomeno quantomeccanico detto effetto tunnel. Sebbene la probabilità che un singolo elettrone attraversi la barriera sia piccolissima, il grande numero di elettroni contenuti nell'arseniuro di gallio drogato comporta che un considerevole numero di elettroni la oltrepassi per effetto tunnel. Se fra le due regioni di arseniuro di gallio drogato non è applicata alcuna differenza di potenziale, il numero degli elettroni che vanno in un senso per effetto tunnel è in media esattamente uguale al numero di quelli che si muovono nel senso opposto; perciò la corrente risultante è nulla. Se invece fra le regioni situate ai due lati della barriera viene stabilita una differenza di potenziale (cioè se a una delle regioni di arseniuro di gallio drogato viene conferita un'energia potenziale maggiore che all'altra), gli elettroni che per effetto tunnel passano dalla regione a potenziale superiore verso quella a potenziale inferiore sono di gran lunga più numerosi di quelli che compiono il passaggio inverso; si genera quindi un flusso di corrente.

La probabilità che gli elettroni subiscano l'effetto tunnel dipende in parte dalla tensione applicata e dallo spessore della barriera. L'energia totale degli elettroni che subiscono l'effetto non varia, perché si tratta di un processo elastico: perciò quando gli elettroni passano da una regione di energia potenziale elevata a una regione di energia potenziale più bassa, la loro energia cinetica aumenta ed essi vengono iniettati a forte velocità. Inoltre gli elettroni che subiscono l'effetto tunnel sono in prevalenza quelli che incidono sulla barriera con velocità più elevate e con direzione più prossima alla perpendicolare. Perciò gli elettroni iniettati costituiscono un fascio ad alta energia che in massima parte punta direttamente in avanti (è perpendicolare al piano della barriera). Inoltre l'energia cinetica con la quale gli elettroni emergono dipende linearmente dalla tensione applicata e quindi è facile da regolare. La barriera a tunnel è quindi particolarmente idonea per i dispositivi a elettroni balistici.

Un dispositivo recente in cui un iniettore di elettroni balistici è particolarmente utile è il cosiddetto transistore verticale. Nella maggior parte dei transistori in commercio le regioni della sorgente e del pozzo sono incastonate in un elemento orizzontale di semiconduttore e gli elettroni che passano dalla prima al secondo costituiscono un flusso parallelo alla superficie del semiconduttore. In un transistore verticale i singoli elementi del dispositivo sono in realtà strati di semiconduttore fatti crescere uno sopra l'altro.

L'eterogiunzione, che è la superficie di separazione fra due tipi di semiconduttore (in questo caso arseniuro di gallio e arseniuro di gallio e alluminio) fatti crescere insieme, consente di iniettare elettroni veloci in uno dei due semiconduttori. Gli elettroni di conduzione situati nell'arseniuro di gallio e alluminio hanno un'energia potenziale superiore a quelli situati nell'arseniuro di gallio. (La curvatura del diagramma dei livelli energetici di ciascun semiconduttore è dovuta alla vicinanza dell'altro semiconduttore.) Quando è applicata una tensione, la funzione energia potenziale varia leggermente (*linea tratteggiata*) e gli elettroni si muovono dall'arseniuro di gallio e alluminio all'arseniuro di gallio, conservando la loro energia totale. Pertanto la loro energia potenziale decresce e quella cinetica aumenta, ed essi si muovono a velocità più elevata (*in alto, in colore*). Invece gli elettroni contenuti nell'arseniuro di gallio hanno un'energia insufficiente per penetrare nell'arseniuro di gallio e alluminio e sono respinti all'interfaccia (*al centro, in colore*).

Un transistore a effetto di campo verticale può essere costruito con la forma di uno stretto piedistallo verticale. Nella parte inferiore vi è un elettrodo, situato direttamente sotto uno strato di semiconduttore che funge da pozzo. Gli strati che costituiscono gli altri elementi, il canale e la sorgente, vengono fatti crescere uno alla volta sopra la regione del pozzo. Un iniettore di elettroni balistici con barriera tunnel, per esempio, può essere costruito semplicemente facendo crescere uno strato conduttore di arseniuro di gallio sopra uno strato di arseniuro di gallio e alluminio (la sorgente), che a sua volta è stato fatto crescere sopra uno strato conduttore di arseniuro di gallio (il canale). Il transistore a effetto di campo verticale presenta anche il vantaggio che la lunghezza del canale conduttore può essere regolata con grande precisione: basta variare lo spessore dello strato di semiconduttore che costituisce il canale. In un transistore orizzontale il canale può essere accorciato solo incastonando nel substrato, in modo tale che risultino vicinissimi tra loro, sorgente, porta e pozzo, un processo che è molto più difficile da controllare con precisione che non quello di costruire uno strato sottile di semiconduttore. (Tuttavia, anche il processo di deporre il metallo della porta sui fianchi di un piedistallo verticale è molto difficile.)

Il primo dispositivo semiconduttore a elettroni caldi (un dispositivo che sfrutta elettroni che hanno subito molti urti, ma possiedono energia e velocità elevate) fu costruito nel 1979 da John Shannon presso i laboratori di ricerca della Philips, in Inghilterra. In questo dispositivo, fatto di silicio drogato, gli elettroni venivano lanciati nel canale conduttore da un iniettore con barriera a gobba di cammello. Dopo aver attraversato il canale, essi dovevano superare una seconda barriera a gobba di cammello prima di raggiungere il collettore. Lo scopo della seconda barriera era quello di impedire agli elettroni termici a bassa energia del canale (che nei dispositivi a elettroni caldi viene di solito chiamato base) di raggiungere il collettore. La seconda barriera era assai più bassa di quella dell'iniettore di elettroni, per consentire ad alcuni elettroni di passare anche se avevano perduto energia negli urti. Alcuni elettroni, iniettati con energia abbastanza elevata, arrivavano al collettore; tuttavia avevano perduto una notevole quantità di energia perché il cammino libero medio degli elettroni nel silicio è corto.

Il transistore di Shannon fu seguito da un transistore di arseniuro di gallio realizzato nel 1981 da Roger Malik, Marc Hollis e Susan Palmateer presso la Cornell University. Il loro dispositivo aveva due barriere a piano drogato con berillio: una barriera alta, che fungeva da iniettore di elettroni caldi, alla sorgente (che lanciava gli elettroni nel dispositivo alla massima velocità possibile) e una barriera più bassa presso il pozzo. Come

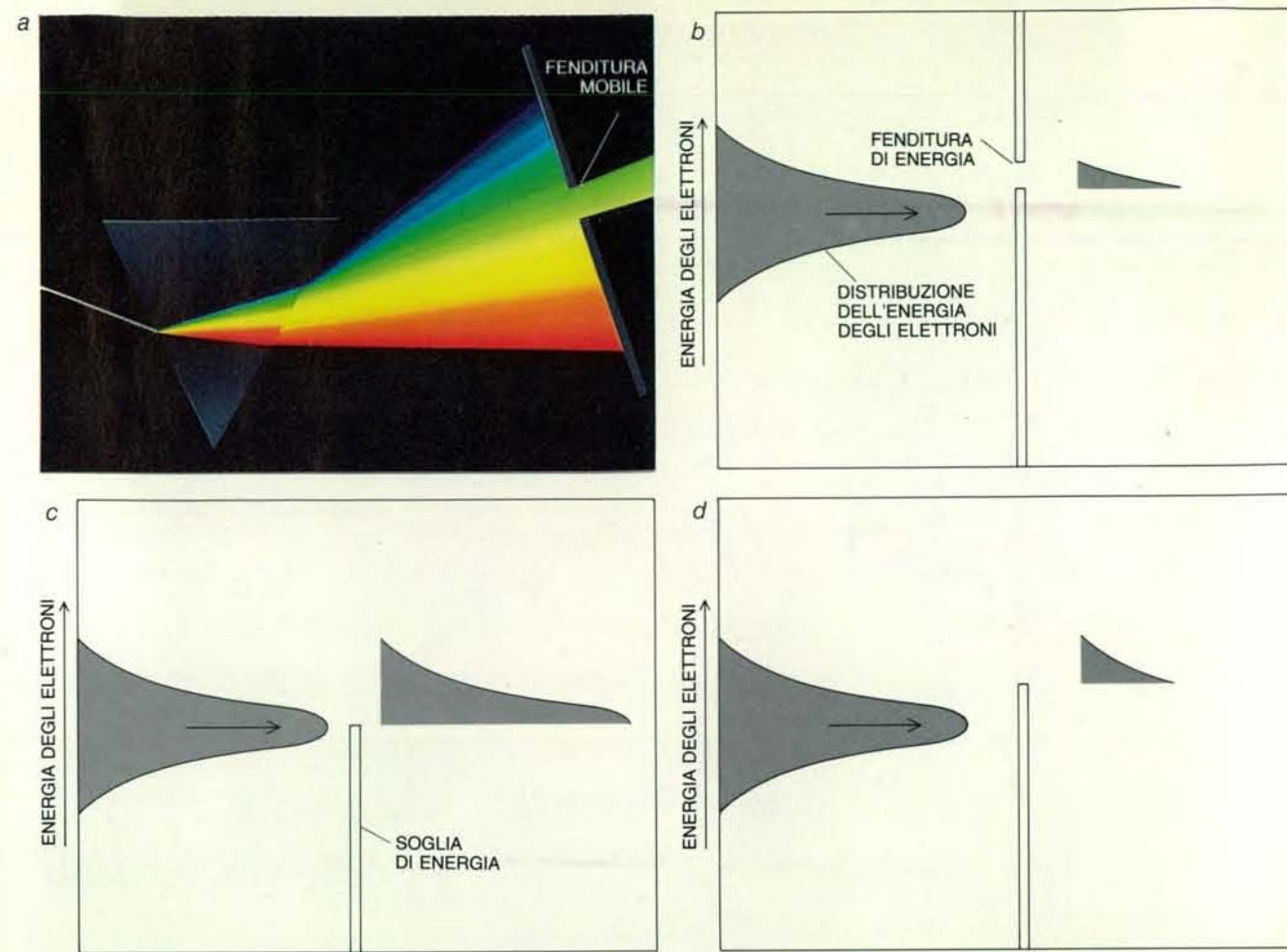
Due tipi di iniettori di elettroni balistici (detti barriera a piano drogato e barriera a effetto tunnel) lanciano elettroni caldi in un semiconduttore. Una barriera a piano drogato viene costruita drogando un campione di arseniuro di gallio con un piano di atomi di berillio (*in alto, a*). Il complesso viene inserito fra due regioni di arseniuro di gallio drogate in modo da essere ricche di elettroni. Il piano di cariche negative in corrispondenza degli atomi di berillio rappresenta una barriera di energia che gli elettroni non superano. Applicando una certa tensione a una delle regioni ricche di elettroni (*b*), l'energia degli elettroni della regione aumenta ed essi valicano la barriera. Quando arrivano dall'altra parte, la loro elevata energia potenziale si trasforma in energia cinetica. Alcuni elettroni subiscono urti dopo aver superato la barriera e si dice che hanno un moto quasibalistico. Una barriera a tunnel (*in basso*) viene costruita inserendo un sottile strato di arseniuro di gallio e alluminio fra due regioni di arseniuro di gallio drogato ricche di elettroni. Gli elettroni dell'arseniuro di gallio non possono entrare nell'arseniuro di gallio e alluminio ma vi è una piccola probabilità quantomeccanica che lo attraversino per effetto tunnel. Se la differenza di potenziale fra le due regioni di arseniuro di gallio è regolata in modo che gli elettroni di una regione abbiano un'energia potenziale maggiore di quelli dell'altra, questi elettroni attraversano la barriera per effetto tunnel, con velocità che dipende dalla differenza di potenziale.

Shannon, anche questi ricercatori osservarono il trasporto degli elettroni caldi, ma non si occuparono in particolare del trasporto balistico.

Fino a tempi molto recenti è stato difficilissimo dimostrare sperimentalmente che in un dispositivo ha avuto luogo trasporto balistico o determinare sperimentalmente la percentuale di elettroni che ha attraversato un dispositivo in modo balistico. A questo scopo occorrerebbe un dispositivo che determinasse la distribuzione dell'energia (per misurare gli eventi di diffusione anelastica) e la direzione del moto (per misurare gli eventi di diffusione elastica) degli elettroni dopo che hanno compiuto l'attraversamento del transistore e le confrontasse con l'energia e la direzione iniziali degli elettroni al momento del loro ingresso nel dispositivo.

La spettroscopia è la misurazione della distribuzione di energia in un fascio di particelle o di radiazione. Un modo per eseguire misure spettroscopiche su un fascio di elettroni sta nel collocare lungo il suo cammino una «soglia di energia». Una soglia di energia lascia passare tutti gli elettroni con energia superiore a un certo valore e respinge tutti gli elettroni che hanno energia inferiore a quel valore. Modificando gradualmente l'«altezza» della soglia (cioè l'energia minima sufficiente per superarla) e misurando la variazione del numero degli elettroni che passano, si può ricavare la concentrazione relativa, nel fascio, degli elettroni aventi energie diverse.

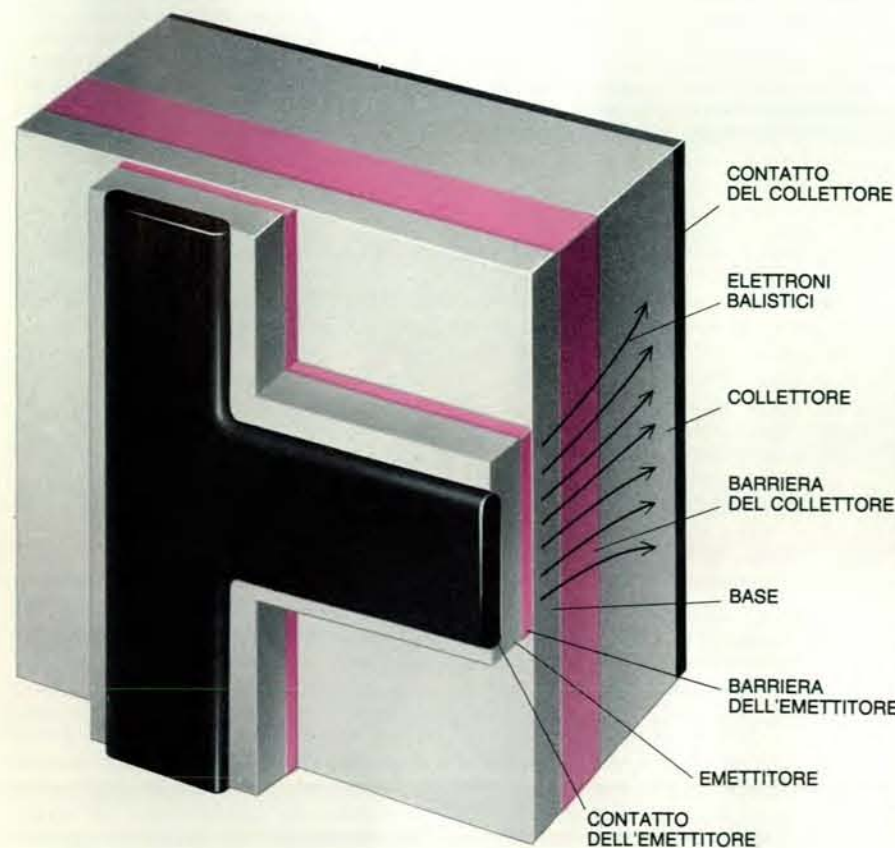
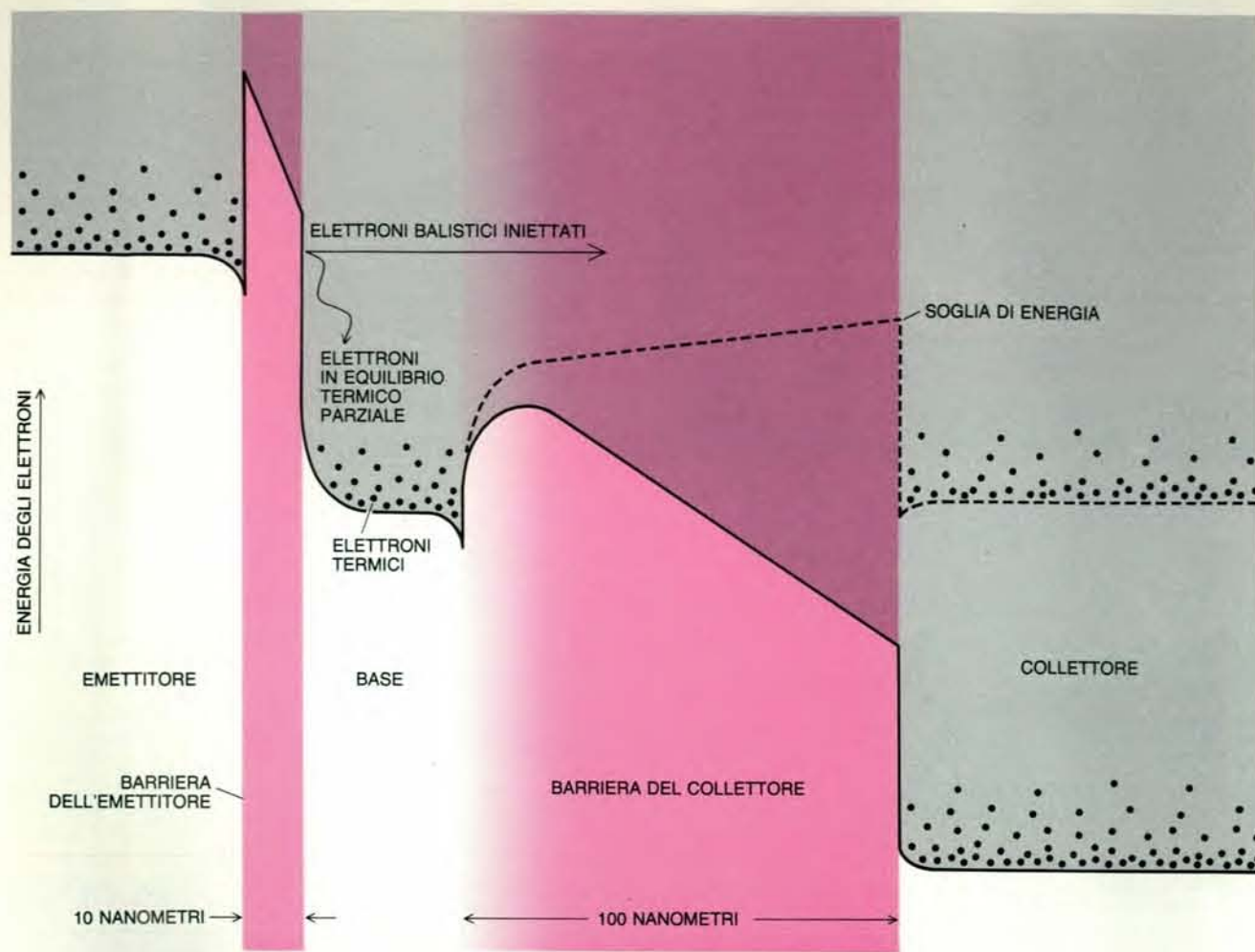
Una semplice soglia di energia è una piastra conduttrice con carica negativa, che respinge gli elettroni. Solo gli elettroni con un'energia abbastanza alta da vincere la forza repulsiva (cioè da superare la barriera di potenziale, di altezza pari alla tensione applicata) possono entrare nella piastra e contribuire alla corrente misurata. L'altezza della soglia di energia può essere cambiata semplicemente variando la tensione applicata alla piastra.

Come si può costruire uno spettrometro che sia sensibile anche alla direzione in cui si muovono gli elettroni, oltre che alla loro energia totale? Un semplice esempio ci permette di immaginarlo. Supponiamo di sparare un proiettile contro una spessa parete verticale. Il proiettile attraversa la parete se ha un'energia superiore a un certo valore minimo. Se tuttavia durante il tragitto viene deviato e non viaggia più in direzione esattamente perpendicolare al muro, dovrà compiere un percorso più lungo all'interno della parete; è come se, in effetti, dovesse attraversare una parete più spessa e perciò dovrà possedere un'energia più elevata. Una parete spessa e verticale misura dunque qualcosa di più specifico dell'energia totale del proiettile: misura l'entità dell'energia che il proiettile cede nel suo moto in direzione perpendicolare alla parete.

Uno spettrometro misura la distribuzione dell'energia in un fascio di radiazione o di particelle. Il primo spettrometro, ideato da Isaac Newton, fu un prisma (*a*). Esso scompone la luce bianca nei colori componenti; ogni colore comprende fotoni aventi una certa energia. Una fenditura mobile consente di far passare e di osservare un solo colore alla volta. Spostando la fenditura si può determinare l'intensità di ciascun colore e quindi la relativa proporzione di fotoni che hanno una data energia. Un dispositivo analogo per compiere una misura spettroscopica su un fascio di elettroni potrebbe consistere in una «fenditura di energia» che facesse passare solo gli elettroni di energia compresa in un certo intervallo (*b*). E' più semplice allestire una «soglia di energia», che fa passare solo gli elettroni aventi un'energia superiore a un certo livello (*c*). Alzando la soglia (*d*) e osservando la variazione del numero di elettroni che passano, è possibile rilevare la composizione energetica del fascio.

In un materiale semiconduttore i proiettili sono sostituiti dagli elettroni e la parete è sostituita da una spessa barriera di energia potenziale (che potrebbe essere una variante più spessa della barriera a tunnel che si trova in certi iniettori di elettroni caldi). L'altezza della barriera di energia potenziale rappresenta la minima energia che occorre a un elettrone per superare la barriera, con una traiettoria perpendicolare a essa. Se la barriera è spessa, la probabilità di attraversamento per effetto tunnel è piccola e può essere trascurata. La distribuzione di energia associata con il moto in avanti degli elettroni in arrivo può essere rilevata modificando l'altezza della barriera e contando il numero degli elettroni che riescono ad attraversare la barriera e danno un contributo alla corrente.

Un dispositivo proposto nel 1981 da uno di noi (Heiblum) è particolarmente adatto a questo genere di spettroscopia perché si basa su un fascio di elettroni collimato che presenta una distribuzione di energia ben conosciuta. Eseguendo la misura spettroscopica in un dispositivo siffatto il ricercatore è in grado di compiere un confronto accurato fra le condizioni iniziali e finali del fascio elettronico; ciò permette di determinare direttamente qualsiasi perdita di energia o qualsiasi cambiamento di direzione siano intervenuti durante il passaggio degli elettroni attraverso il dispositivo. Esso è noto con il nome di THETA (tunneling hot-electron-transfer amplifier).

I dispositivi THETA costruiti finora sono per lo più fatti di arseniuro di gallio e di arseniuro di gallio e alluminio. Sono stati costruiti da un gruppo giapponese (presso i laboratori Fujitsu) e da Heiblum e collaboratori (presso il Thomas J. Watson Research Center della IBM). Nel dispositivo THETA una barriera a effetto tunnel inietta elettroni a velocità elevatissima in un sottile strato conduttore chiamato base. La base è costituita da arseniuro di gallio reso conduttore grazie a un drogaggio con atomi di silicio che forniscono gli elettroni «freddi», cioè elettroni in equilibrio termico con il reticolo, i quali consentono di applicare una differenza di potenziale ai capi della barriera a tunnel. L'energia degli elettroni iniettati ha una distribuzione abbastanza stretta intorno a un valore centrale che è determinato dalla tensione applicata all'iniettore a effetto tunnel.

Lo spessore della base è minore oppure dello stesso ordine di grandezza del cammino libero medio degli elettroni caldi e quindi molti di essi l'attraversano in modo balistico. All'altra estremità di questa sottile base vi è una spessa barriera costituita da uno strato di arseniuro di gallio e alluminio. La barriera impedisce agli elettroni freddi di uscire dalla base per effetto tunnel, ma può essere superata dagli elettroni caldi che hanno attraversato la base in modo balistico.

Un dispositivo THETA (è l'acronimo di tunneling hot-electron-transfer amplifier) genera e rivela elettroni balistici. Il diagramma energetico del dispositivo (*in alto*) ne illustra il funzionamento. Gli elettroni dell'emettitore sono mantenuti a un'energia elevata. Alcuni di questi attraversano per effetto tunnel una barriera, per propagarsi poi con moto balistico attraverso il canale conduttore (chiamato base). Alcuni di questi elettroni non sono balistici: vengono diffusi e diventano termici (tendono all'energia di equilibrio degli elettroni di conduzione presenti nella base). La barriera del collettore impedisce a tutti gli elettroni non balistici di entrare nel collettore. Tale barriera può essere usata come uno spettrometro: regolando la tensione del collettore si crea una soglia di energia. È così possibile determinare la composizione energetica degli elettroni che si propagano nel dispositivo e confrontarla con la distribuzione di energia che possiedono quando vengono lanciati dall'emettitore.

Dall'altra parte della barriera un altro strato conduttore di arseniuro di gallio funge da collettore. In esso gli elettroni perdono la maggior parte della loro energia cinetica a causa di vari meccanismi di diffusione e vengono attratti verso l'elettrodo positivo della batteria (*si veda l'illustrazione nella pagina a fronte*).

Per usare un dispositivo THETA come spettrometro a soglia, si iniettano nella base sottile elettroni caldi con una distribuzione di energia nota, mentre si aumenta progressivamente l'altezza della barriera di potenziale spessa. L'altezza della barriera viene aumentata applicando una tensione negativa (che respinge gli elettroni) al collettore. La soglia dello spettrometro è costituita dal lato dello strato spesso di arseniuro di gallio e alluminio che è più vicino al collettore, poiché qui la barriera di energia potenziale è più alta e poiché solo gli elettroni con un'energia maggiore o uguale all'altezza della barriera riescono ad attraversare tutto lo strato di arseniuro di gallio e alluminio e a giungere al collettore.

Via via che l'altezza della barriera viene aumentata, la corrente che giunge al collettore decresce. La variazione di corrente provocata da ogni variazione dell'altezza della barriera è proporzionale al numero di elettroni che hanno un'energia (in realtà si tratta dell'energia associata al moto perpendicolare alla barriera) maggiore di quella richiesta per superare la barriera prima che la sua altezza venga modificata, ma minore dell'energia necessaria per superare la barriera dopo che la sua altezza è stata modificata. Quindi variando l'altezza della barriera è possibile rilevare la distribuzione di energia degli elettroni che hanno attraversato il dispositivo.

In parecchi esperimenti di questo tipo il numero e gli spettri degli elettroni giunti sul collettore si sono rivelati in ottimo accordo con le caratteristiche del fascio iniettato. I risultati spettroscopici costituiscono una dimostrazione diretta che gli elettroni si sono spostati con moto balistico nello strato conduttore di arseniuro di gallio e attraverso la barriera spessa di arseniuro di gallio e alluminio. In questi esperimenti è stata osservata per la prima volta una così stretta distribuzione di energia degli elettroni che hanno attraversato un dispositivo a semiconduttore; è stata anche la prima volta che le caratteristiche degli elettroni emergenti si sono dimostrate tanto simili a quelle degli elettroni iniettati.

In futuro i dispositivi THETA e altri dispositivi a elettroni balistici potranno forse trovare applicazione nei componenti a semiconduttori per il mercato; tuttavia essi hanno già dimostrato di essere utili strumenti scientifici per studiare il comportamento degli elettroni nei semiconduttori. I meccanismi implicati nel passaggio di un elettrone attraverso un semiconduttore non sono ancora ben compresi e i dispositivi a elettroni balistici forniscono ai ricercatori la possibilità di modificare alcune condizioni all'interno del semiconduttore e di misurare l'effetto di queste variazioni sul passaggio degli elettroni.

Per esempio nei dispositivi THETA, non tutti gli elettroni iniettati attraversano la base con moto balistico. Con apparente paradosso, tuttavia, se si rende meno conduttrice la base riducendo il numero di atomi dell'agente drogante, il numero degli elettroni balistici aumenta. Questo risultato indica che gli elettroni caldi che attraversano la base interagiscono con gli elettroni freddi che la rendono conduttrice (e forse anche con gli atomi droganti). Non è ancora chiaro di che natura sia questa interazione.

Consideriamo un altro esempio. Se la base del dispositivo THETA viene ispessita, la distribuzione di energia degli elettroni balistici raccolti dal collettore non muta di forma, benché il loro numero complessivo decresca. Ciò dimostra che ogni evento di diffusione riguarda in modo casuale un solo elettrone caldo sottraendolo completamente alla distribuzione degli elettroni balistici. In altre parole, non vi sono eventi di diffusione importanti di piccolo angolo o anelastici che comportino una perdita di energia molto piccola. Ogni collisione è «traumatica» e raffredda un elettrone caldo.

Un'altra area d'indagine in cui gli elettroni balistici possono fungere da preziose sonde sperimentali è lo studio delle proprietà quantomeccaniche fondamentali dell'elettrone. Quando vengono diffusi, in particolare quando vengono diffusi in modo anelastico, gli elettroni perdono in parte il «ricordo» delle loro precedenti proprietà ondulatorie. In particolare perdono l'informazione relativa alla fase dell'onda. Gli elettroni balistici, d'altra parte, si propagano in un mezzo solido senza essere diffusi e conservano la loro natura ondulatoria. Per gli elettroni balistici è dunque possibile ripercorrere all'indietro, nel tempo e nello spazio, il moto dell'onda elettronica. La relazione fra due elettroni balistici rimane costante fino a quando essi rimangono balistici.

Il comportamento ondulatorio degli elettroni balistici nei semiconduttori può essere compreso grazie all'analogia con le onde luminose e con le onde del mare. Quando un'onda di questo genere colpisce un parete, talvolta viene riflessa: viene cioè prodotta un'altra onda che si propaga all'indietro verso la sorgente dell'onda originaria. L'onda incidente e l'onda di ritorno (riflessa) interferiscono fra loro. Quando una cresta della prima incontra una cresta della seconda, l'ampiezza complessiva aumenta: questo è il fenomeno dell'interferenza costruttiva. Analogamente dove il ventre di un'onda incontra la cresta dell'altra, l'ampiezza complessiva diminuisce: questo è il fenomeno dell'interferenza distruttiva. Passando da un mezzo all'altro (per esempio dall'aria al vetro) spesso le onde luminose vengono parzialmente riflesse: in parte la luce attraversa la superficie di separazione fra i due mezzi e in parte viene riflessa, formando un'onda che può interferire con il raggio incidente. Analogamente quando gli elettroni balistici passano dall'arseniuro di gallio all'arseniuro di gallio e alluminio o viceversa, parte dell'onda elettronica attraversa la barriera e parte viene riflessa verso la sorgente degli elettroni.

Nel dispositivo THETA si può evidenziare l'interferenza fra l'onda di elettroni incidente e quella di ritorno. Dopo essere entrati nella base sottile e averla attraversata, gli elettroni balistici subiscono riflessioni quantomeccaniche sulla superficie di separazione tra base e collettore. Dato che la lunghezza d'onda di un'onda elettronica dipende dalla sua energia, gli effetti d'interferenza causati dalle riflessioni dipendono dall'energia degli elettroni iniettati. Per certi valori dell'energia di iniezione le onde elettroniche riflesse tornano all'iniettore a effetto tunnel in modo da essere esattamente in fase con le onde che escono per la prima volta dall'iniettore (cioè in modo che vi sia una sovrapposizione perfetta). Allora l'onda di ritorno rinforza quella incidente interferendo costruttivamente con essa e ciò fa affluire nel collettore una maggiore quantità di corrente. Per altri valori dell'energia le onde interferiscono distruttivamente e l'onda di ritorno sopprime in parte quella incidente. Allora la corrente che va al collettore diminuisce. Al variare dell'energia degli elettroni iniettati, questi effetti quantistici danno luogo a massimi e minimi locali facilmente osservabili nella corrente che attraversa il dispositivo.

E' sorprendente, e in contrasto con la nostra intuizione, che nei cristalli purissimi, a temperature prossime allo zero assoluto, gli elettroni possano teoricamente percorrere grandi distanze senza essere diffusi dalla densa matrice degli atomi. Sotto certi aspetti, tuttavia, è ancora più incredibile che gli elettroni possano muoversi senza subire diffusione anche nei dispositivi reali, che sono fatti di materiali impuri e funzionano a temperature molto superiori allo zero assoluto.

La dimostrazione sperimentale che gli elettroni possano muoversi con moto balistico è importante sotto diversi aspetti. Come si è visto, la natura ondulatoria degli elettroni balistici fornisce ai ricercatori nuovi strumenti per esplorare il comportamento degli elettroni nei solidi e per verificare la teoria quantomeccanica dell'elettrone. Sotto un profilo più pratico, gli elettroni balistici nei semiconduttori consentono di progettare nuovi dispositivi quantistici che potrebbero funzionare a velocità molto superiori a quelle consentite ai dispositivi elettronici della generazione attuale.

I nuovi *Lagerstätten* fossili del Triassico italiano

Le ricerche condotte su numerosi depositi altamente fossiliferi scoperti in questi ultimi anni nell'Italia settentrionale consentono di ampliare notevolmente le conoscenze paleontologiche sugli organismi del Triassico

di Giovanni Pinna

Nel 1970 un paleontologo dell'Università di Tübingen, A. Seilacher, indicò con il nome di *Fossil-Lagerstätten*, un termine che significa semplicemente «giacimenti di fossili», alcuni particolari accumuli paleontologici. Più precisamente con questo termine Seilacher indicò «corpi sedimentari che forniscono una inusuale quantità di informazioni paleontologiche» sia in senso qualitativo, sia in senso quantitativo, e cioè giacimenti nei quali il grande numero di fossili, la loro varietà o la perfezione della loro conservazione forniscono, oltre a un grande numero di dati, precise informazioni circa i meccanismi della fossilizzazione, l'ambiente di sedimentazione e l'ambiente di vita degli organismi rinvenuti.

L'inusuale quantità e qualità delle informazioni paleontologiche contenute nei Fossil-Lagerstätten permette cioè, secondo Seilacher, di analizzare a fondo tutti gli eventi che hanno dato origine all'accumulo paleontologico e che sono responsabili della conservazione dei resti fossili, dalla distruzione delle parti molli e del materiale scheletrico degli organismi (necrolisi), al trasporto sedimentario e al seppellimento dei resti organici (biostratinomia), alla distruzione degli scheletri mineralizzati (diagenesi), fino alla ricostruzione delle condizioni del biotopo originario (paleobiologia).

Fra i Fossil-Lagerstätten Seilacher incluse sia giacimenti celebri per la conservazione del tutto eccezionale degli organismi fossili - quelli che conservano tracce delle parti molli e che in varie occasioni io ho definito come «giacimenti a conservazione totale» - sia accumuli fossili estremamente ricchi di esemplari di vario tipo, come i calcari a crinoidi o i *bone bed* (brecce ossifere).

Fra i primi, certamente i più spettacolari, possiamo qui ricordare a titolo di esempio i giacimenti giurassici di Holzmaden e di Solnhofen in Germania, il giacimento giurassico di Osteno in Lombardia (si veda il mio articolo *I fossili giurassici di Osteno* in «Le Scienze» n. 193, settembre 1984) e il giacimento cambriano di Burgess Shales nella Columbia Britannica (si veda l'articolo *La fauna degli argilloscisti di Burgess* di S. Conway Morris e H. B. Whittington in «Le Scienze» n. 133, settembre 1979), nei quali molti organismi si sono conservati con tracce delle parti molli e con strutture molto delicate, o gli accumuli di ambra di varie parti del mondo con inclusi numerosi invertebrati, soprattutto insetti, conservati in toto.

La varietà e la perfezione dei fossili contenuti distingue nettamente i Lagerstätten fossili dalle normali rocce fossilifere che, seppure a volte ricche di esemplari fossili, forniscono, tuttavia indicazioni assai scarse ai fini della ricostruzione degli eventi del passato

I Lagerstätten fossili costituiscono dunque la fonte primaria di informazioni paleontologiche; essi si sono formati (e probabilmente si formano ancora oggi) in condizioni ambientali (biologiche e fisiche) del tutto particolari. Sebbene le combinazioni di fattori biotici e abiotici che producono un giacimento paleontologico siano innumerevoli, vi sono tuttavia solo poche situazioni che possono portare alla formazione di Lagerstätten fossili, i quali perciò sono tutt'altro che abbondanti. Secondo Seilacher questi si formano per concentrazione o per conservazione di organismi. La differenza fra i due tipi di giacimento paleontologico è abbastanza evidente. I depositi di concentrazione sono giacimenti nei quali organismi che di solito si trovano sparsi nei sedimenti, con un numero di esemplari molto limitato, vengono riuniti soprattutto da fattori fisici. Ciò che rende importante un deposito di concentrazione non è perciò tanto la qualità dei fossili o la loro rarità (poiché essi si trovano anche nelle normali rocce fossilifere) quanto il loro alto numero e la grande varietà di tipi diversi.

Fra i depositi di concentrazione si possono includere le fessure sottomarine, nelle quali si concentrano spesso i resti di numerosi organismi, i depositi alluvionali e le grotte.

I depositi di conservazione sono invece quelli responsabili, come si è già accennato, di fossilizzazioni eccezionali, sono cioè giacimenti che, grazie a particolari condizioni dell'ambiente di deposizione, hanno conservato organismi, impronte o tracce di attività che in altre rocce vanno solitamente distrutti. Ciò che rende importanti questi giacimenti non è quindi la quantità dei fossili, ma la loro qualità e, spesso, la loro unicità.

È possibile distinguere essenzialmente tre tipi diversi di depositi di conservazione: i depositi di stagnazione, i depositi di soffocamento e le trappole di conservazione.

I depositi di stagnazione sono certamente i più diffusi; sono depositi che si formano in linea generale in bacini lacustri o in bacini marini ristretti, con scarsa ossigenazione, privi di forti correnti e di abbondante sedimentazione clastica, bacini nei quali le tracce (quando esistono) e gli organismi, trasportati da aree più ossigenate quali gli strati d'acqua più superficiali o le zone circostanti a ossigenazione normale, si sono conservati in modo perfetto con le strutture più delicate grazie alla mancanza di agenti distruttori meccanici e biologici.

Depositi di stagnazione sono quelli che abbiamo citato di Holzmaden e di Solnhofen, il giacimento triassico di Besano in Lombardia e quello cretacico del Libano.

I depositi di soffocamento sono invece depositi che si formano per rapido infossamento dovuto a sedimentazione abbondante. È il caso delle torbiditi, frane sottomarine costituite da materiale molto fine, o delle tempestiti, nubi di sedimenti trasportate da tempeste e temporali e risedimentate. Il fango che inquina improvvisamente l'ambiente subacqueo provoca il soffocamento degli organismi che non sono in grado di fuggire (non i pesci, per esempio, ma gli echinodermi che hanno inoltre un delicato sistema respiratorio), il loro seppellimento e, quindi, la loro conservazione. È attraverso la successione di frane sottomarine che si formò nel Cambriano il giacimento di Burgess Shales e nel Devoniano il giacimento tedesco di Bundenbach; gli organismi contenuti in quest'ultima località sono infatti soprattutto echinodermi, stelle di mare e crinoidi, morti soffocati.

Le trappole di conservazione, infine, sono giacimenti statisticamente più rari che si formano a seguito della presenza di trappole naturali che imprigionano gli organismi e li conservano grazie alla presenza di condizioni particolarmente favorevoli alla fossilizzazione (mancanza di ossigeno ecc.). Trappole di conservazione sono per esempio le torbiere e i giacimenti di ambra.

Da tutto ciò che si è detto deriva che la scoperta di nuovi Lagerstätten fossili assume sempre una grande importanza in paleontologia, poiché crea la possibilità di un notevole incremento nella conoscenza degli eventi del passa-

I crostacei abbondano nel calcare di Zorzino. La maggior parte di essi, circa il 90 per cento, è attribuibile a specie nectoniche, la più frequente delle quali è *Antrimpos noricus* (*in alto a sinistra*). Esemplari di questa specie si rinvengono in tutte le località fossilifere, assai spesso sotto forma di exuvie. Meno frequenti sono i crostacei bentonici, quali *Protoclytiopsis dubia* (*in alto a destra*) e *Archaeopalinurus levis* (*in basso a sinistra*), la più antica aragosta conosciuta. Più rari sono i rappresentanti della classe tilacocefali, attribuiti alla specie *Microcaris minuta* (*in basso a destra*); di essi è conservato solo il carapace cefalico. Queste, come le altre fotografie dell'articolo, sono dell'autore e di Luciano Spezia.

Nella ricostruzione paleogeografica della Tetide durante il Triassico superiore (circa 200 milioni di anni fa) sono indicate le facies principali: le terre emerse (*in marrone*), i mari epicontinentali (*in blu chiaro*), le piattaforme carbonatiche (*in blu scuro*) e le facies marine profonde (*in grigio*).

to. Queste scoperte non sono però frequenti. Nonostante questa rarità dei Lagerstätten, negli ultimi anni sono venuti alla luce nell'Italia settentrionale numerosi giacimenti di fossili del Triassico superiore che possono essere classificati come tali.

Nel 1972 in una cava situata presso l'abitato di Cene in Val Seriana, nelle Prealpi bergamasche, aprentesi in una serie di calcari scuri regolarmente stratificati, si verificò una grande frana: parte degli strati inclinati nel senso del pendio scivolarono verso la base della cava, mettendo allo scoperto per un largo tratto uno strato di pochi centimetri di spessore estremamente ricco di fossili ben conservati, soprattutto pesci e crostacei.

Le ricerche effettuate nella cava di Cene dal Museo di scienze naturali di Bergamo portarono al rinvenimento di un grande numero di fossili: centinaia di pesci, molte decine di crostacei e alcuni rettili volanti, i più antichi rappresentanti di questo gruppo animale oggi estinto.

Il calcare che affiora nella cava di Cene fa parte di una formazione geologica ritenuta fino a poco tempo fa priva o assai povera di fossili: la formazione del calcare di Zorzino, costituita da una serie di calcari micritici (ossia a grana molto fine) in strati di spessore variabile da qualche centimetro a circa mezzo metro, spesso separati da interstrati argillosi e intercalati a dolomie e a marne nerastre. I calcari della formazione di Zorzino affiorano in diverse località dell'Italia settentrionale e sono associati alle dolomie del Triassico superiore note con il nome di dolomia principale. Essi affiorano in varie zone del Bergamasco, delle Prealpi bresciane e della Provincia di Udine (ove prendono il nome di dolomie di Forni), e mantengono in tutte queste località una notevole costanza litologica.

Il rinvenimento di fossili abbondanti e perfettamente conservati nel calcare di Zorzino affiorante a Cene fu l'inizio di una serie di ricerche su tutti gli affioramenti di questa formazione: tali ricerche portarono al rinvenimento di orizzonti fossiliferi in numerose località. Il Museo di scienze naturali di Bergamo, che aveva già scavato a Cene, e l'Istituto di paleontologia dell'Università di Milano rinvennero livelli fossiliferi in altre località del Bergamasco, in Val Brembilla e presso i paesi di Zogno e di Endenna; livelli fossiliferi furono rinvenuti e sfruttati nel Bresciano, in Val Garza e in Val Vestino, dai Musei di storia naturale di Brescia e di Milano, mentre ricerche furono effettuate in Val Preone, fra Tolmezzo e Ampezzo, dal Museo friulano di storia naturale.

Il risultato di queste ricerche è stato il ritrovamento di una fauna fossile molto ampia e di estremo interesse, una fauna che allarga notevolmente le conoscenze paleontologiche sui crostacei, sui pesci e sui rettili in un epoca, il Triassico superiore appunto, nel quale questi gruppi animali erano in una fase delicata del loro sviluppo evolutivo.

I fossili rinvenuti nel calcare di Zorzino sono tuttavia particolarmente interessanti anche per un'altra ragione: grazie alla loro varietà, alla loro perfetta conservazione, al fatto che sono spesso completi e molto abbondanti si può tentare infatti la ricostruzione ambientale di una vasta area di sedimentazione di mare poco profondo che durante il Triassico superiore, circa 200 milioni di anni fa, era compresa fra i margini dell'Eurasia e dell'Africa, un'area in cui si depositarono imponenti spessori di sedimenti carbonatici che sono oggi trasformati in rocce calcaree e dolomitiche relativamente povere di fossili, fra cui la dolomia principale cui prima ho fatto cenno.

Da questo punto di vista i giacimenti paleontologici del calcare di Zorzino possono perciò essere considerati dei veri e propri Lagerstätten fossili.

Si è detto che il calcare di Zorzino è sempre associato alla dolomia principale: esso costituisce infatti una formazione differenziata, ma coeva (eteropica) in facies bituminosa di detta dolomia. Ciò significa che mentre la dolomia principale si sedimentava in alcuni luoghi sotto certe condizioni ambientali, in altri luoghi, caratterizzati da altre condizioni ambientali, si depositavano calcari diversi ricchi di materia organica.

La dolomia principale, chiamata dai geologi di lingua tedesca *Hauptdolomit*, è una formazione nel complesso poco fossilifera costituita da dolomie cristalline, alternate a dolomie bioclastiche, a dolomie micritiche e a dolomie stromatolitiche, che si formò per sedimentazione calcarea soprattutto durante il Norico, e cioè durante la parte mediana del Triassico superiore. La dolomia principale fa parte di un'importante massa di rocce calcaree triassiche che affiora con continuità da Gibilterra alla Turchia. Si ritiene che queste rocce calcaree, e quindi anche la dolomia principale, si siano formate per sedimentazione calcarea su piattaforme carbonatiche molto estese, analoghe a quelle che oggi si trovano nelle Bahamas e nel Golfo Persico, e cioè su bassofondi marini di profondità non superiore ai 10-15 metri che potevano periodicamente affiorare formando vaste pianure di marea o terre emerse che venivano esposte all'alterazione subaerea.

Durante il Triassico superiore queste piattaforme carbonatiche erano molto diffuse nella regione occidentale della Tetide, nell'area cioè in cui questo oceano si insinuava fra il margine meridionale del continente eurasiatico e il margine settentrionale dell'Africa, formando un ampio golfo. In particolare questi bassofondi a sedimentazione calcarea seguivano tutti i margini degli antichi continenti triassici sui quali predominava un ambiente arido e all'interno dei quali si insinuavano a volte mari epicontinentali poco profondi soggetti a forte evaporazione. All'esterno delle piattaforme carbonatiche, verso est, si stendeva l'oceano Tetide, ove si depositavano calcari di facies più profonda, mentre alcuni margini delle piattaforme erano limitati, verso l'oceano, da barriere organogene.

I movimenti verticali del livello del mare producevano su queste vaste piattaforme un certo numero di condizioni ambientali diverse, mutevoli nel tempo anche localmente. Così, mentre nelle aree più lontane dai margini della piattaforma, caratterizzate da mare meno profondo e soggette a trasformarsi in pianure di marea o in basse aree emerse, si depositava la dolomia principale, nelle aree più spostate verso i margini oceanici della piattaforma si formava una roccia leggermente differente che prende il nome di «calcare di Dachstein». Essa rappresenta condizioni di mare poco più profondo e passa a volte a calcari organogeni di scogliera.

Nella dolomia principale, che affiora con molta continuità lungo tutte le Alpi meridionali e che raggiunge spessori notevoli nel Bergamasco e nel Bresciano, sono state identificate varie fasi di sedimentazione che si alternano con regolarità, formando veri e propri cicli sedimentari. Si è osservato in particolare che ogni ciclo è composto dalla successione di una fase sopratidale erosiva costituita da una semplice superficie di discontinuità o da un livello argilloso, di una fase intertidale trasgressiva composta soprattutto da dolomie stromatolitiche e micritiche, di una fase subtidale costituita da dolomie cristalline massicce con macrofossili e di una fase intertidale regressiva, analoga alla fase trasgressiva. Ogni ciclo si ripeté più volte nel corso della sedimentazione della dolomia principale, il che significa che la piattaforma carbonatica era soggetta a cicliche variazioni di profondità: a una fase di emersione seguiva una trasgressione marina, la piattaforma diveniva quindi una pianura di marea e poi un fondale marino poco profondo; una successiva regressione del mare portava nuovamente la piattaforma a livello di pianura di marea e poi all'emersione, e così via. Durante le fasi sopratidali, la piattaforma emergeva per ampi tratti, nelle fasi intertidali essa dava origine a vaste pianure con estese superfici algali stromatolitiche e nelle fasi subtidali non superava probabilmente i 10 o 15 metri di profondità.

Come si è accennato, all'interno della dolomia principale, in più località e a livelli stratigrafici diversi, sono comprese formazioni eteropiche calcareo-bituminose, sono cioè compresi i calcari di Zorzino (o facies analoghe): queste formazioni non costituiscono un *continuum*, ma sono frazionate in bacini di sedimentazione separati che, nelle Prealpi bergamasche, corrispondono alla porzione sommitale della dolomia principale, nelle Prealpi bresciane alla parte mediana e nelle Alpi venete ai livelli inferiori.

Si ritiene che queste formazioni calcareo-bituminose siano il risultato della sedimentazione in bacini ristretti che si formavano, soprattutto per cause tettoniche, all'interno della piattaforma carbonatica. E cioè, in alcuni punti della piattaforma, a causa della diversa velocità di subsidenza delle varie parti della piattaforma stessa, si creavano bacini interni, più profondi, subsidenti, attivi lungo scarpate di faglia, sul fondo dei quali si alternavano fasi di sedimentazione tranquilla, con decantazione di fanghi micritici e di argille, e fasi di maggior apporto clastico, dovuto allo scivolamento lungo i fianchi dei bacini e alla deposizione di torbiditi.

Le dimensioni di questi bacini erano molto varie. Nelle Prealpi bresciane, si può stimare per tre diversi bacini un'ampiezza variabile da tre a cinque chilometri, con una potenza di sedimenti accu-

In tutte le località fossilifere del calcare di Zorzino i pesci sono molto abbondanti e ben conservati. Lo studio del materiale estratto fino a ora non è ancora terminato e numerose forme, quali il pesce volante (*in alto*), non sono ancora state descritte. Tuttavia fra il materiale studiato è da segnalare la presenza di numerose specie appartenenti a gruppi sistematici molto diversi. I selaci sono rappresentati da serie isolate di denti di una sola specie, *Pseudodalatias barnstonensis*. Fra gli osteitti sono presenti sia i crossotterigi con rari celacanti, sia i condrostei, con i generi *Saurichthys*, *Birgeria*, *Thoracopterus*, *Pholidopleurus* e *Olophagus*, sia gli olostei con rappresentanti dei semionotiformi, amiiformi, picnodontiformi e folidoforiformi. Questi ultimi sono i pesci di gran lunga più abbondanti (*in basso*) presenti nei giacimenti con almeno tre specie diverse: *Parapholidophorus nybelini*, *Pholidoctenus serianus* e *Pholidorhinchodon malzanii*.

mulati variabile da 250 a 600 metri. Esistono tuttavia bacini di dimensioni anche maggiori: nel Bergamasco è stato recentemente studiato un bacino di 12-15 chilometri di larghezza che doveva raggiungere una profondità di 260 metri ed era circondato da pendii con una inclinazione media di soli quattro gradi.

Secondo tutti i ricercatori che si sono occupati di queste formazioni, i bacini che si aprivano nella piattaforma carbonatica erano bacini ristretti, poco profondi, con acque stagnanti al fondo, privi di ossigeno (anossici) e probabilmente sovrasalati. Tutto ciò si deduce dalle caratteristiche del sedimento e del suo contenuto organico, e soprattutto dall'assenza di fauna bentonica autoctona e di infauna, dalla persistenza di fini laminazioni non algali negli strati e dall'abbondanza di materia organica diffusa.

In generale i bacini di questo tipo sono depositi di conservazione: sono cioè bacini nei quali la vita presso i fondali è impossibile per la mancanza di ossigeno dovuta alla stagnazione dell'acqua e all'accumulo di materia organica, ma nei quali gli organismi, caduti dopo la morte sul fondo, hanno la possibilità di conservarsi perfettamente, per mancanza di agenti distruttori meccanici e biologici, e cioè per l'assenza di correnti, di predatori e di batteri decompositori. Si ritiene generalmente che questi fossili ben conservati, di solito pesci, crostacei, molluschi ed echinodermi non bentonici, e rettili marini, indichino che gli strati d'acqua più superficiali dei bacini erano normalmente ossigenati e quindi abitati da una fauna nectonica normale, mentre i margini dei bacini, anch'essi a ossigenazione normale, possedevano una normale fauna bentonica. I fossili contenuti in questi depositi dovrebbero perciò rappresentare quel che resta degli organismi nectonici viventi negli strati d'acqua più superficiali e degli organismi bentonici che vivevano ai margini del bacino e che potevano cadere o essere trasportati dopo la morte sul fondo del bacino stesso, ove, grazie alle particolari condizioni ivi esistenti, si sono conservati.

Fra i rettili rinvenuti nei livelli fossiliferi del calcare di Zorzino vi sono diversi pterosauri che risultano essere i più antichi rettili volanti conosciuti. Nella cava di Cene sono stati trovati sei esemplari e un dente isolato che sono stati attribuiti a due diverse specie: *Eudimorphodon ranzii* (*nella fotografia*) e *Peteinosaurus zambellii*. Una terza specie, *Preonodactylus buffarinii*, è stata descritta su un esemplare proveniente da Val Preone (Udine); a questa specie va attribuito anche un frammento d'ala trovato a Endenna. Poiché si ritiene che i rettili volanti, nel complesso cattivi volatori, dovessero spiccare il volo partendo da alte rocce costiere o dalle cime degli alberi, la presenza di questi animali sembra dimostrare l'esistenza di terre emerse ricche di vegetazione.

Questa caduta e conservazione dei resti organici rendono evidente l'importanza dei Lagerstätten fossili nella ricostruzione dell'ambiente di sedimentazione. Nel caso specifico di un deposito di conservazione la fauna contenuta permette infatti di effettuare una ricostruzione ambientale non solo dei bacini di conservazione stessi, ma anche delle aree circostanti.

Anche per quanto riguarda il calcare di Zorzino si può ritenere che i fossili, contenuti in alcuni ben delimitati livelli fossiliferi, possano permettere la ricostruzione ambientale non solo dei bacini anossici in cui sono conservati, ma anche della piattaforma carbonatica circostante, e cioè dell'area di deposizione della dolomia principale, sulla quale i dati paleontologici non sono particolarmente abbondanti. Per far ciò è però necessario che ogni resto organico fossilizzato, ogni eventuale impronta organica o sedimentaria presente vengano considerati come il frutto di un evento eccezionale, meritevole della più grande attenzione per essere compreso.

La dolomia principale è, come si è detto, complessivamente povera di fossili: essa contiene livelli ad alghe stromatolitiche (veri e propri tappeti algali) diffusi nelle fasi intertidali, alghe calcaree e molluschi diffusi nelle fasi subtidali. Proprio gli strati formatisi nelle fasi subtidali, quando cioè la piattaforma carbonatica era sommersa, presentano uno spettro faunistico molto povero. Questi sedimenti contengono infatti relativamente poche specie di lamellibranchi, soprattutto megalodontidi, pochi gasteropodi, qualche brachiopodo e qualche celenterato. La povertà di macrofossili della dolomia principale viene messa in relazione con la distruzione dei resti organici che sarebbe stata operata dalla dolomitizzazione di un originale sedimento calcareo. Tuttavia, nonostante la possibile cancellazione di molte tracce organiche, l'associazione faunistica mostra ancora una diversità tassonomica bassa, associata a un alto numero di esemplari per ogni taxon. Questi due caratteri (evidenti soprattutto nei banchi a megalodontidi) sarebbero indicativi di un ambiente di sedimentazione prossimo a condizioni estreme, probabilmente per salinità e temperatura elevate.

L'analisi della fauna contenuta nei livelli fossiliferi del calcare di Zorzino ci fornisce uno spettro faunistico più completo e complesso. I fossili che si rinvengono in questa formazione sono classificabili sulla base del loro habitat originale in tre gruppi: organismi marini bentonici (originariamente viventi sulla piattaforma carbonatica), organismi marini nectonici (viventi sia sulla piattaforma carbonatica, sia negli strati d'acqua sovrastanti i bacini anossici), organismi continentali (viventi sulle terre emerse che presumibilmente si formavano durante le fasi di emersione della piattaforma).

Gli organismi marini bentonici sono molto scarsi: mancano assolutamente gli animali endobionti, mentre gli epibionti sono rappresentati da pochi echinodermi (soprattutto ofiuroidi, pochissimi asteroidi ed echinoidi), da qualche briozoo e qualche celenterato, da crostacei e da scarsi molluschi. I lamellibranchi in particolare sono tutti di piccole dimensioni e appartengono a specie caratterizzate da guscio leggero, già segnalate nella dolomia principale, quali *Isognomon exilis*, *Myoconcha cornalbae*, *Eopecten egidii* e *Macrodon rudis*.

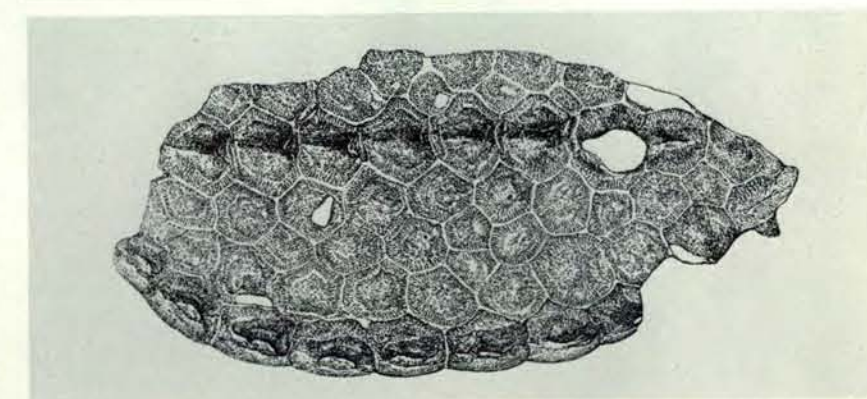
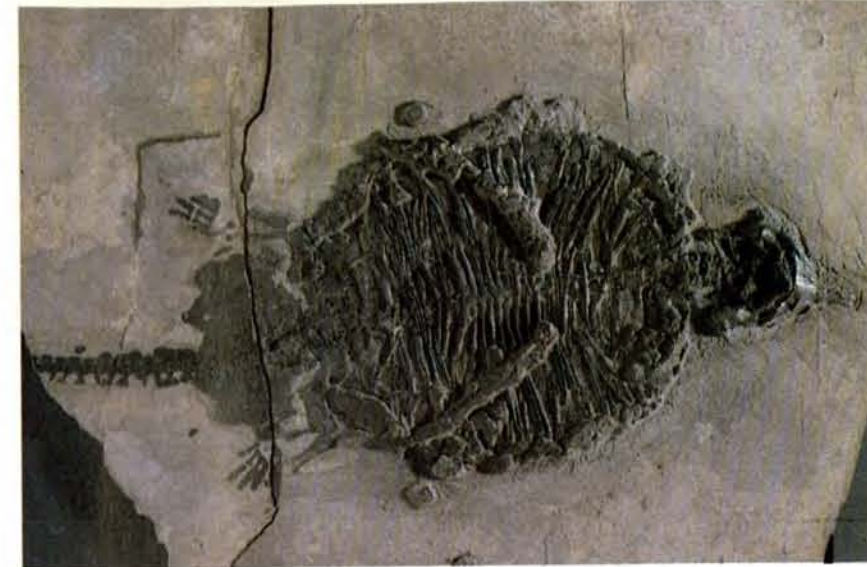
Per quanto concerne i crostacei è interessante notare che le forme bentoniche rappresentano solo il 10 per cento dell'intera fauna. Si tratta di astacidei (*Protoclytiopsis dubia*), di palinuridi (*Archaeopalinurus levis*), di un peneideo (*Palaeodusa longipes*) e di un rappresentante della classe dei tilacocefali (*Microcaris minuta*).

Più abbondanti sono gli organismi nectonici. Fra questi vi è il 90 per cento dei crostacei, attribuibile a peneidei (*Antrimpos noricus*, *Aeger* sp.) e a caridei. La maggior parte di questi crostacei è di piccole dimensioni (da due a tre centimetri di lunghezza) ed è rappresentata soprattutto da exuvie, cioè da spoglie abbandonate dagli animali con la muta.

I pesci sono molto abbondanti; la maggior parte è costituita da forme di piccole dimensioni (folidoforiformi) e da specie la cui vita era legata al substrato (come i picnodontiformi). Fra i pesci, e in particolare fra i folidoforiformi, molti esemplari sono più o meno vistosamente incurvati, ciò che avviene quando una spoglia organica è soggetta a disidratazione sia per esposizione all'aria, sia a causa dell'alta salinità dell'acqua.

Al necton vanno infine riferiti alcuni rettili. Si tratta dei placodonti, rettili dotati di una corazza e di una dentatura adatta a una dieta a base di molluschi o, comunque, di organismi dotati di conchiglia, di cui sono stati rinvenuti diversi esemplari a Endenna, riferibili a una specie prossima a *Psephoderma alpinum*. I placodonti erano organismi poco mobili, la cui vita si svolgeva soprattutto presso i fondali.

Anche il lepidosauro *Endennasaurus acutirostris*, di cui è stato rinvenuto a Endenna un solo esemplare, era un animale marino, sebbene fosse legato, più dei placodonti, all'ambiente costiero.

Gli organismi continentali rinvenuti nei bacini anossici sono rettili e vegetali. Fra i rettili vi sono tre specie di rettili volanti (i più antichi rappresentanti del gruppo degli pterosauri), due tecodonti (*Mystriosuchus* sp. e *Megalancosaurus*

I rettili placodonti sono relativamente abbondanti nel giacimento di Endenna. Finora sono stati rinvenuti un dente isolato, il cranio di un esemplare molto giovane, un frammento di corazza (*in basso*), una corazza dorsale completa, un esemplare molto incompleto e un esemplare completo (*in alto e al centro*). Tutti questi resti sono attribuibili a una specie prossima allo *Psephoderma alpinum* del Retico della Baviera e della Lombardia. I placodonti sono un gruppo di rettili marini triassici dotati di denti larghi e piatti adatti a masticare organismi con conchiglia o esoscheletro resistente, e cioè molluschi, brachiopodi e crostacei. La specie rinvenuta a Endenna è dotata di una corazza larga e appiattita, di una coda lunga ma bassa e di arti posteriori ridotti. Queste caratteristiche inducono a pensare che questo animale non fosse un buon nuotatore, ma che fosse invece adattato alla vita su bassi fondali fangosi. È probabile perciò che questo placodonte poco mobile abitasse la piattaforma carbonatica, avesse la capacità di infossarsi nel fango calcareo del fondo e si nutrisse di piccoli lamellibranchi. Il muso lungo e appuntito sembra infatti interpretabile come una struttura adatta alla raccolta di lamellibranchi endobionti. (Il disegno dello scheletro completo è di Massimo Demma, quello della corazza è di Annalisa Durante.)

preonensis), un primitivo lepidosauro scavatore (*Drepanosaurus unguicaudatus*) e alcune altre forme non ancora pubblicate (*Vallesaurus cenensis* e *Macrocnemus* sp.).

Tracce di vegetazione presumibilmente continentale si rinvengono nei livelli fossiliferi, ma la pessima conservazione di tali resti non permette una adeguata analisi della flora.

È da segnalare infine la presenza nei livelli fossiliferi di un'alta percentuale di escrementi fossili (coproliti), riferibili a pesci e a crostacei, e di ammassi interpretabili come rigurgiti o contenuti gastrici.

In tutti i bacini analizzati questa fauna è concentrata in livelli rocciosi laminati separati da pacchi di strati privi o quasi di fossili. Dai dati in mio possesso si ricava che nella cava di Cene il livello fossilifero da cui provengono in gran numero crostacei, pesci e rettili è spesso circa sei-sette centimetri. A Endenna sono stati segnalati in una sequenza di tre metri di potenza tre livelli principali a pesci di 37, 56 e 35 centimetri di spessore rispettivamente, separati da strati privi di fossili.

La presenza di livelli ricchi di fossili in una serie sedimentaria è spesso considerata come la prova dell'esistenza nei luoghi di deposito o, come nel nostro caso, nei luoghi circostanti, di condizioni favorevoli allo sviluppo della vita. Questo però non corrisponde sempre a verità; il più delle volte infatti questi livelli molto fossiliferi indicano condizioni di vita così estreme da causare, per un minimo spostamento dei fattori ambientali, la morte in massa di organismi di ogni tipo. Ciò è vero soprattutto se, come avviene nel calcare di Zorzino, i livelli fossiliferi sono discontinui, sono cioè separati da intervalli rocciosi privi di fossili. Questa discontinuità dei livelli fossiliferi deve infatti corrispondere a una discontinuità degli eventi che li hanno prodotti, il che significa che solo in alcuni momenti sopravvennero particolari condizioni che causarono una concentrazione di spoglie organiche che avevano la possibilità di fossilizzarsi.

L'alta concentrazione dei fossili in limitati livelli fossiliferi sembra escludere inoltre la possibilità che il deposito si sia formato per normale sedimentazione di spoglie organiche cadute verticalmente dagli strati d'acqua superiori o dai margini dei bacini normalmente ossigenati. La morte di organismi, quali pesci e crostacei, e la deposizione delle loro spoglie organiche sul fondo in condizioni normali sono infatti in un ambiente marino eventi così diluiti nel tempo e nello spazio da non potere dar luogo a particolari concentrazioni.

È chiaro perciò che i livelli fossiliferi discontinui contenuti nel calcare di Zorzino si formarono a seguito di eventi eccezionali, a loro volta discontinui; ed è chiaro altresì che i fossili conservati in tali livelli rappresentano perciò la fauna originaria della piattaforma carbonatica, trasportata da questi eventi eccezionali all'interno dei bacini anossici ove ebbe la possibilità di conservarsi.

Come documentano la mancanza di bioturbazioni, la persistenza di fini laminazioni negli strati e l'accumulo di materia organica, sul fondo dei bacini ove si depositò il calcare di Zorzino vi erano condizioni tali da impedire lo sviluppo della vita. La stagnazione dell'acqua al fondo di questi bacini, dovuta al loro isolamento, e l'accumulo della materia organica producevano condizioni anossiche che, a loro volta, impedendo la presenza di decompositori e di predatori, hanno permesso la perfetta conservazione dei resti organici.

Anche gli scarsi organismi bentonici rinvenuti nei livelli fossiliferi non sono perciò originari dei bacini anossici, ma dovevano provenire dalla piattaforma carbonatica. Ciò che colpisce è l'estrema scarsità di questi organismi bentonici, che indica che anche l'ambiente di piattaforma non doveva essere particolarmente favorevole. La stessa indicazione ci viene fornita dagli organismi nectonici: questi, seppure più abbondanti dei bentonici, presentano pur sempre una bassa diversità tassonomica, da cui si può dedurre che le condizioni esistenti sui fondali della piattaforma carbonatica non erano tali da permettere un abbondante sviluppo della fauna e della flora.

Il quadro generale che ne deriva è che si trattasse di un basso fondale a fanghiglia calcarea, ad alta salinità, ben illuminato, come dimostra la presenza di alghe calcaree, con scarsa vegetazione e con radi e discontinui banchi a molluschi e accumuli di alghe calcaree e briozoi, un basso fondale lontano dal mare aperto sul quale raramente si avventuravano le specie di grandi dimensioni (mancano per fare un esempio gli ittiotterigi) e sul quale viveva invece una fauna relativamente monotona di piccoli pesci e di piccoli crostacei, assieme ad alcuni rettili, i placodonti, strettamente legati alla vita sul fondo marino.

Il *Drepanosaurus unguicaudatus* è il rettile più inusuale fra quelli rinvenuti nel calcare di Zorzino. Di lui si conoscono un esemplare completo, ma privo del cranio (*in alto a sinistra*), e tre esemplari giovanili incompleti provenienti due da Endenna e uno da Val Preone. Il drepanosauro è una lucertola primitiva, lunga circa 45 centimetri, le cui più strane caratteristiche, ben visibili nel disegno qui a fianco, sono un artiglio potente ed enormemente sviluppato nel secondo dito dell'arto anteriore e un robusto uncino situato all'estremità della coda. Questi due caratteri, assieme allo sviluppo di alcuni elementi del cinto scapolare e delle neurapofisi delle prime vertebre dorsali che servivano ad assicurare l'attacco di potenti muscoli, inducono a ritenere che egli fosse un rettile scavatore, che si procurasse cioè il cibo scavando, come fa oggi il formichiere o l'oritteropo. Un animale di questo tipo non poteva dunque che vivere in ambiente francamente terrestre. (I disegni sono di Massimo Demma.)

Il *Megalancosaurus preonensis*, di cui è qui illustrato un particolare, è un rettile del gruppo dei tecodonti (lo stesso che dette origine sempre nel Triassico ai rettili volanti, ai dinosauri e ai coccodrilli). Se ne conosce un solo esemplare rinvenuto in Val Preone. Di questo animale, non ancora studiato in dettaglio, si sa ben poco. Si ritiene che fosse un rettile arboricolo.

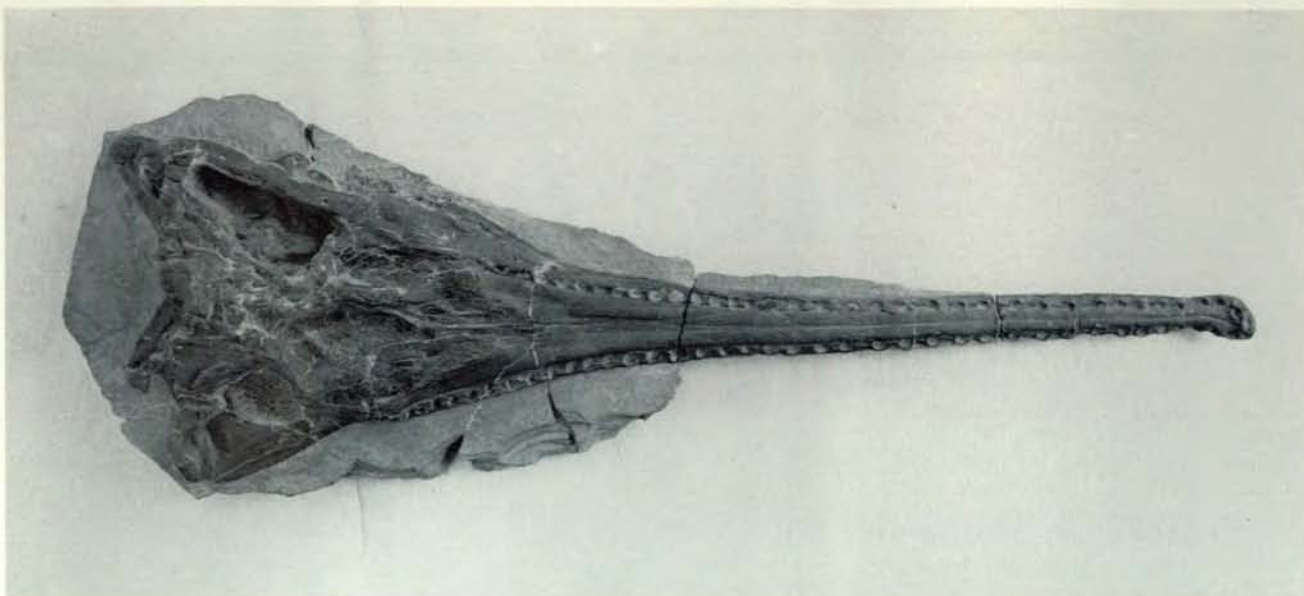
Livelli fossiliferi a pesci e a crostacei sono presenti in tutti gli affioramenti noti del calcare di Zorzino; più in generale livelli fossiliferi analoghi a quelli rinvenuti nel calcare di Zorzino sono presenti in tutte le facies bituminose eteropiche ai calcari e alle dolomie di piattaforma del Triassico superiore. Ciò significa, da un lato, che all'interno delle piattaforme carbonatiche vi era una costante presenza di bacini più profondi, di limitata estensione e, almeno in alcuni momenti, a circolazione d'acqua ristretta; dall'altro, che gli eventi paleontologici responsabili della formazione dei livelli fossiliferi ebbero luogo più volte in modo analogo in varie località e dovevano essere perciò collegati in qualche modo all'interazione fra l'ambiente di piattaforma e quelli dei bacini.

Non è facile individuare quale meccanismo favorì la concentrazione delle spoglie organiche in precisi livelli stratigrafici durante la sedimentazione del calcare di Zorzino. Tuttavia è possibile azzardare alcune ipotesi del tutto generali.

La discontinuità dei livelli fossiliferi può indicare sostanzialmente due cose: in primo luogo che anche gli eventi che portarono all'accumulo delle spoglie organiche furono discontinui, ebbero cioè luogo in alcuni intervalli temporali separati da intervalli privi di eventi paleontologici evidenti; in secondo luogo, che solo in alcuni intervalli temporali si instaurarono al fondo dei bacini le condizioni adatte alla conservazione delle spoglie organiche.

Le due possibilità conducono a due diverse ricostruzioni dei meccanismi di formazione dei livelli fossiliferi. Nel primo caso si deve ipotizzare che i bacini presenti all'interno della piattaforma fossero molto costanti nelle loro caratteristiche fisiche, e cioè che al fondo di tali bacini vi fossero costanti condizioni di anossicità tali da permettere sempre la conservazione dei resti fossili. In questo caso la formazione delle concentrazioni fossili sarebbe da imputare a eventi esterni ai bacini, e cioè a eventi fisici operanti sulla piattaforma.

Nel secondo caso si deve invece ipotizzare che le condizioni al fondo dei bacini variassero nel tempo; per esempio, che, a seguito di variazioni nella circolazione d'acqua al fondo, si alternassero condizioni anossiche (che permettono la conservazione delle spoglie organiche) e condizioni ossidanti (che permettono lo sviluppo della vita, ma che portano alla distruzione di resti organici delicati come pesci e crostacei). In questo caso la formazione delle concentrazioni fossili non sarebbe perciò da imputare a eventi fisici esterni ai bacini, ma alla variazione delle possibilità di fossilizzazione all'interno dei bacini stessi.

Questa seconda ipotesi non sembra tuttavia avvalorata dai dati in nostro possesso: gli strati che separano i livelli fossiliferi a pesci e a crostacei non conten-

Dal giacimento di Endenna proviene un cranio completo di *Mystriosuchus* (probabilmente *Mystriosuchus planirostris*). Si tratta di un rettile tecodonte del gruppo dei fitosauri, tipico del Triassico superiore europeo. Era un animale anfibio, della taglia di un gaviale di medie dimensioni, adattato alle acque salmastre delle lagune. La sua presenza è dunque legata all'esistenza di lagune e di terre emerse nella regione.

gono infatti tracce organiche evidenti, né normali fossili (per esempio molluschi ed echinodermi) che dovrebbero essere invece presenti se si fossero instaurate, negli intervalli temporali corrispondenti alla deposizione di queste rocce, normali condizioni di vita sul fondo.

Per contro in numerosi livelli privi di fossili della formazione di Zorzino vi sono evidenti tracce di eventi sedimentari: strati gradati dovuti ad accumulo di sedimenti a seguito di torbiditi, scivolamenti e frane sottomarine. Tutto ciò indica che vi erano frequenti movimenti di materiale dai margini verso il fondo dei bacini, e sembra poter avvalorare l'ipotesi che anche l'accumulo dei fossili possa aver avuto luogo a seguito di fenomeni di questo tipo.

La fauna contenuta nei livelli fossiliferi mostra tuttavia alcune particolarità: essa sembra composta soprattutto da organismi mobili, leggeri e quindi facilmente trasportabili, quali crostacei ed exuvie di crostacei, pesci di piccole dimensioni e mascelle di piccoli squali. Mancano invece, o sono molto scarsi, gli organismi poco mobili legati strettamente alla vita sui fondali, quali organismi endobionti o epibionti, quali echinodermi e molluschi. Ciò sembrerebbe indicare che i livelli fossiliferi si siano formati per accumulo sul fondo di materiale organico trasportato dalla piattaforma circostante a seguito di correnti o di tempeste.

L'assenza dai livelli fossiliferi di organismi bentonici poco o affatto mobili (per esempio l'assenza dei megalodontidi che sulla piattaforma vivevano in vasti banchi, semiinfossati nei sedimenti) pare tuttavia contrastare con un trasporto dei resti organici dalla piattaforma ai bacini operato da eventi turbolenti, quali frane sottomarine o tempeste.

I dati parzialmente contrastanti forniti dall'analisi della fauna diverrebbero più comprensibili se si ipotizzasse una correlazione tra la formazione dei livelli fossiliferi e le oscillazioni del livello del mare sulla piattaforma carbonatica, correlazione che appare però assai difficile da effettuare sul piano sedimentologico e stratigrafico.

Si potrebbe pensare cioè che, quando a seguito dell'abbassamento relativo del livello marino porzioni più o meno estese della piattaforma emergevano localmente o si trasformavano in pianure di marea, i bacini aprentisi sulla piattaforma si trasformassero in bacini chiusi, sovrasalati e quasi totalmente anossici. In questi venivano intrappolati gli organismi più mobili fra quelli viventi sulla piattaforma in condizioni di immersione (piccoli pesci, crostacei, rettili marini) e venivano concentrati, attraverso i canali di marea, i resti organici depositati sul fondo fangoso della piattaforma (per esempio le mascelle degli squali e le exuvie dei crostacei).

L'accumulo di organismi in bacini-trappola di questo tipo spiegherebbe l'alta concentrazione sia di organismi mobili in rapporto alla scarsità di forme bentoniche, sia la presenza contemporanea di resti organici, evidentemente depositati sulla piattaforma e trasportati dopo la morte, e di organismi intrappolati nel bacino ancora viventi. Inoltre, in un bacino di accumulo normalmente sommerso, nel quale le spoglie organiche fossero trasportate turbolentemente dalle zone circostanti, gli organismi bentonici dovrebbero essere presenti in una concentrazione ben maggiore, anche se le zone circostanti fossero caratterizzate, come nel caso della piattaforma carbonatica del Triassico superiore, da bassa diversità tassonomica conseguente a condizioni ambientali non ottimali.

La presenza di vertebrati terrestri e di resti di vegetali continentali può avere inoltre un ruolo non indifferente a favore di quest'ultima ipotesi. Questi elementi dimostrano infatti che, durante gli intervalli in cui si formarono i livelli fossiliferi, attorno ai bacini di deposizione dovevano essere presenti terre emerse di una certa estensione, ricche di vegetazione e popolate da una fauna ben diversificata. La distanza di queste terre emerse dai bacini di deposizione non è facilmente valutabile poiché le spoglie dei vertebrati possono galleggiare per lungo tempo grazie allo sviluppo di gas di putrefazione all'interno del corpo. Tuttavia nel nostro caso i vertebrati rinvenuti non sono minimamente decomposti e si presentano con tutte le ossa in connessione anatomica, il che rende poco probabile un flottaggio troppo lungo e quindi distanze troppo grandi fra il luogo di origine e il luogo di deposizione.

La difficoltà nell'interpretare i meccanismi di formazione dei livelli fossiliferi rende problematica la classificazione di questi giacimenti: sulla base delle ipotesi effettuate essi potrebbero infatti essere considerati sia trappole di concentrazione, sia trappole di conservazione. Tuttavia, al di là della loro classificazione, resta il fatto che questi Lagerstätten del Triassico superiore rappresentano una preziosa fonte di nuove informazioni sul passato biologico della Terra.

Matrimonio, maternità e carriera scientifica

La produttività pubblicistica delle donne è inferiore a quella degli uomini, ma gli impegni familiari non bastano a spiegare questa differenza, dato che le donne con marito e figli pubblicano quanto le colleghe nubili

di Jonathan R. Cole e Harriet Zuckerman

Gli studi sul rendimento dei ricercatori scientifici, misurato in base alla loro produttività in termini di pubblicazioni, rivelano che, a parità di età, di istituto in cui è stato conseguito il dottorato e di settore di attività, nel corso di tutta la loro carriera le donne pubblicano meno saggi degli uomini. Per spiegare questa differenza sono state avanzate svariate ipotesi, che vanno dalla sistematica discriminazione sessuale a differenze di natura biologica, finora mai dimostrate, per quel che riguarda la predisposizione a studi di carattere scientifico.

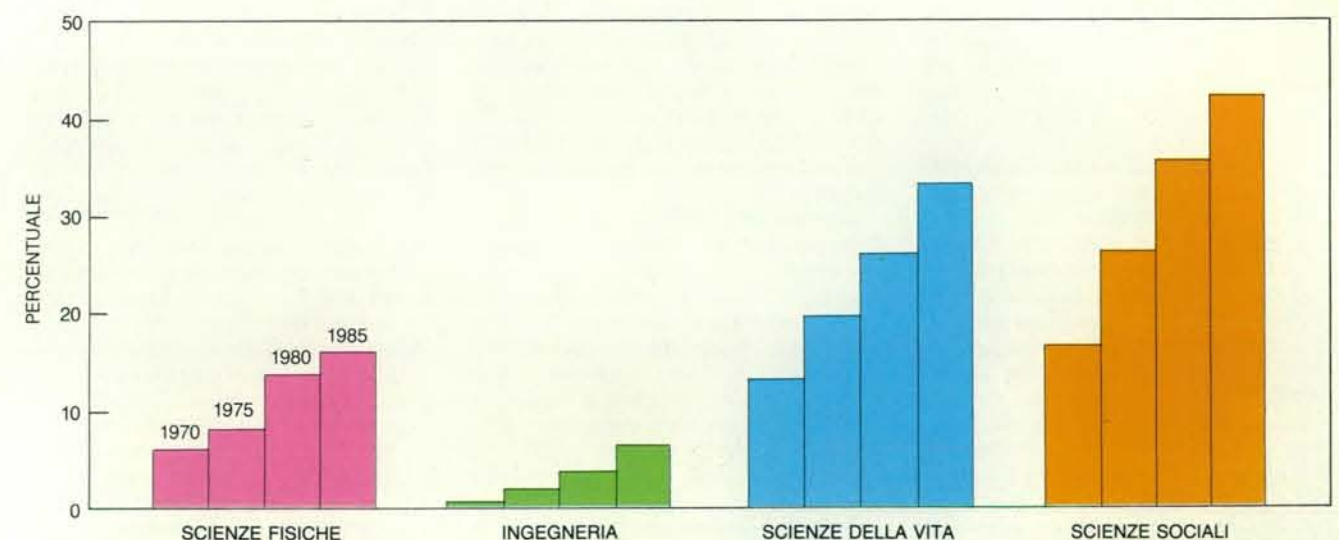
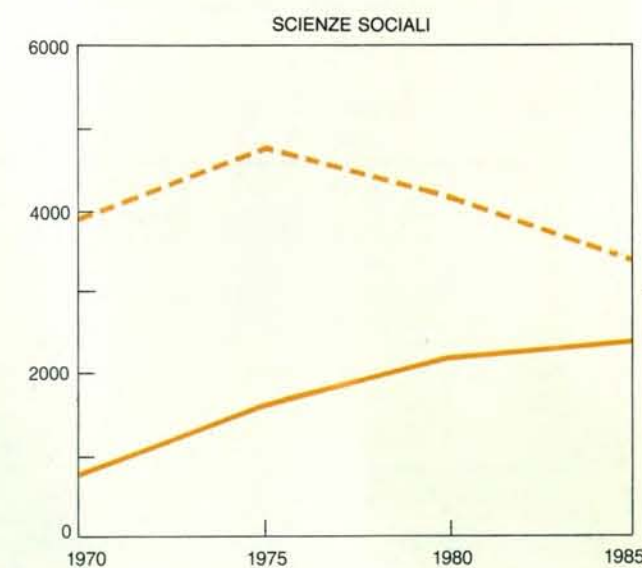
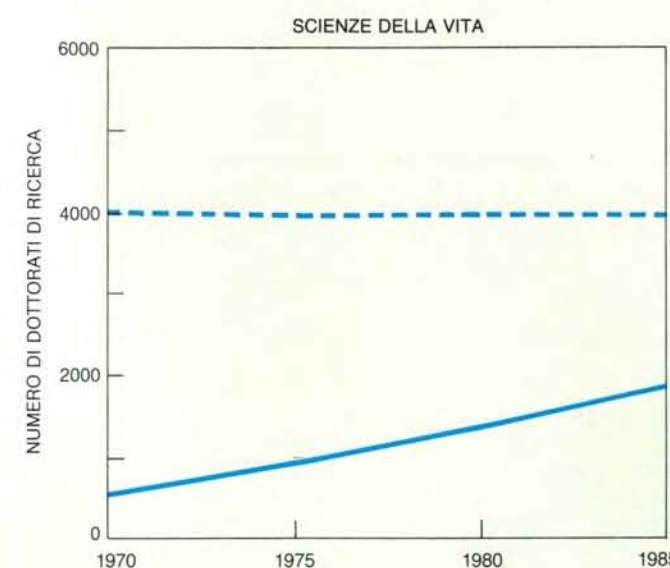
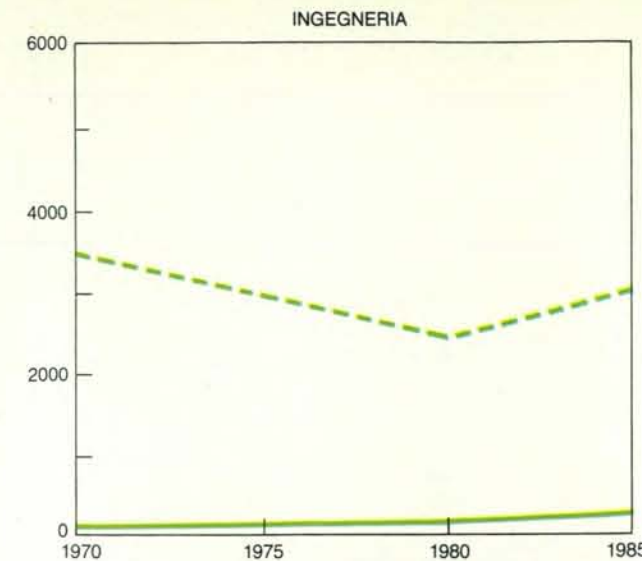
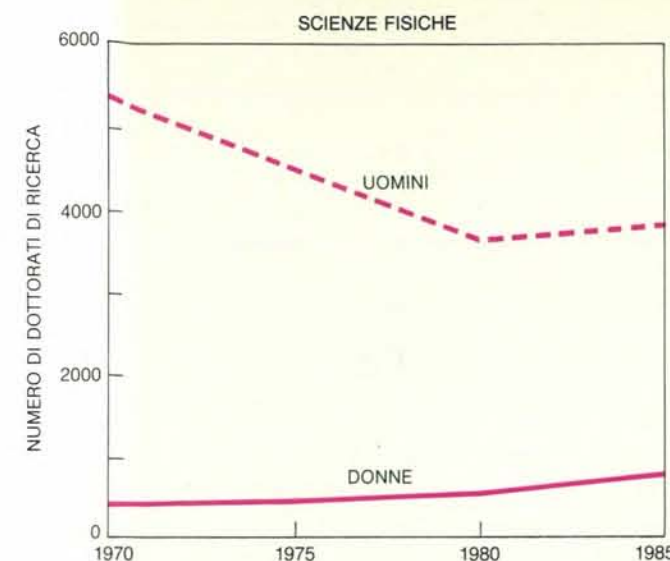
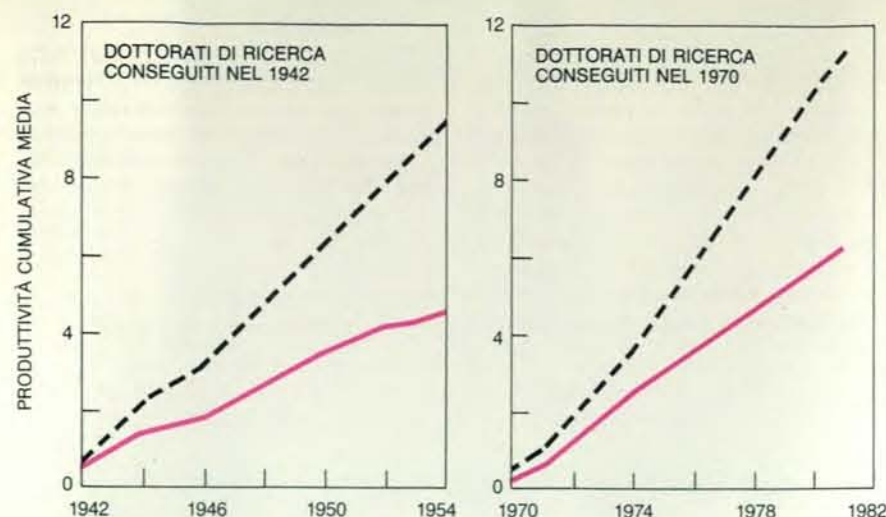
Una spiegazione proposta di frequente vuole che sulle donne, molto più che sugli uomini, gravi il peso del matrimonio e della cura dei figli e che questo ruolo nella vita sociale spieghi nel migliore dei modi le differenze osservate fra i due sessi in materia di pubblicazioni scientifiche. Vera o non vera che sia questa spiegazione, la convinzione che le cose stiano proprio così incide di fatto sulle opportunità di carriera delle donne, sulle loro decisioni e sul trattamento a esse riservato.

Noi abbiamo deciso (come parte di una più vasta indagine sulla carriera delle donne e degli uomini di scienza americani) di verificare la controaffermazione, sostenuta in studi precedenti, secondo la quale matrimonio e maternità non hanno il minimo effetto sulla carriera scientifica delle donne. Lo abbiamo fatto tenendo conto del rapporto dinamico tra vita familiare e lavoro di ricerca delle donne nel corso di tutta la loro carriera: impostazione, questa, mai adottata dai precedenti ricercatori, i quali non facevano altro che correlare il numero di saggi pubblicati dalle donne e il loro attuale ruolo di mogli e di madri. Il nostro studio si basa su colloqui effettuati con 120 scienziati: 73 donne e 47 uomini. Volevamo sapere se gli scienziati (sia uomini sia donne) ritengono che il matrimonio e la paternità o la maternità siano incompatibili con una carriera scientifica in generale, se questo sia stato vero per loro in particolare e quali effetti quantificabili (misurati dal numero di pubblicazioni) il matrimonio e la maternità abbiano avuto di fatto sulla carriera scientifica delle donne. Poiché per tradizione gli uomini non hanno mai avuto la responsabilità primaria della cura dei figli, ci siamo concentrati quasi interamente sulle donne, mettendo a confronto il tasso di pubblicazione delle coniugate e delle nubili, delle madri e di quelle che non hanno figli.

Il calcolo delle pubblicazioni, pur essendo un indicatore impreciso del contributo scientifico dei ricercatori, è tuttavia in stretta correlazione con unità di misura migliori, come la valutazione espressa dalla comunità scientifica nel suo complesso; la prolificità pubblicistica è inoltre di grande importanza per la carriera scientifica. Abbiamo considerato quindi il numero di pubblicazioni come criterio di valutazione approssimativo, ma utile, di quest'ultima. Ci siamo resi conto che le ricercatrici sono, in un certo senso, delle «sopravvissute». Per definizione, esse sono passate attraverso i rigori di tutti gli studi universitari, hanno ottenuto il dottorato di terzo grado e sono riuscite a inserirsi nel lavoro scientifico. Non abbiamo invece preso in esame l'influenza esercitata dalle aspettative culturali sulle probabilità delle donne di superare queste prove: ciò avrebbe reso necessario lo studio dei processi mediante i quali le donne vengono spinte ad abbandonare la carriera scientifica.

Abbiamo scelto i soggetti per questo studio con un processo selettivo stratificato che teneva conto di sesso, anzianità di servizio, settore di competenza e reputazione scientifica. Per confrontare gli effetti del matrimonio e della maternità sulle donne che si sono laureate in periodi storici differenti, abbiamo suddiviso le donne in tre gruppi di età: quelle che hanno conseguito il dottorato tra il 1920 e il 1959, prima dell'avvento del movimento femminista, e quelle che lo hanno conseguito rispettivamente nei decenni 1960-1969 e 1970-1979, quando il movimento in questione si stava mettendo in cammino e poi si diffondeva ampiamente. L'80 per cento delle donne si interessava di matematica e di scienze fisiche e biologiche e il resto di economia e di psicologia; il campione venne scelto in modo che questo stesso rapporto valesse per ciascun gruppo di età.

Gli scienziati in esame sono stati ulteriormente suddivisi in gruppi di soggetti con pari riconoscimenti accademici, rispetto ad altri aventi più o meno la stessa anzianità di servizio. La fascia superiore del gruppo più anziano era costituita da scienziati (quelli da noi definiti «eminenti») che erano membri della National Academy of Sciences o dell'American Academy of Arts and Sciences oppure erano professori ordinari in dipartimenti stimati fra i primi 10 in ogni settore di studio in base a indagini nazionali sulla qualità dei programmi di studio per il conseguimento del dottorato. Nel gruppo di età intermedio sono stati definiti eminenti i borsisti della fondazione Guggenheim o i professori associati in uno dei 10 dipartimenti più prestigiosi. Tra gli scienziati più giovani sono stati considerati eminenti quelli che fruivano di una borsa di studio della fondazione Guggenheim o erano assistenti o professori associati in uno dei 10 dipartimenti più importanti.

Gli scienziati che non rispondevano a questi severi requisiti sono stati definiti «gregari». Essi venivano scelti a caso in elenchi del personale di ricerca e di dipartimento presso scuole superiori e uni-

Dal 1970 il numero dei dottorati di ricerca conferiti a donne è aumentato nonostante una generale flessione del numero di quelli conferiti a uomini. Le curve (*in alto*) mostrano il numero annuo di lauree di terzo grado negli Stati Uniti dal 1970 a tutto il 1985. La percentuale conferita alle donne è aumentata nettamente (*in basso*), ma rimane bassa in fisica e in ingegneria. I dati sono tratti da studi del National Research Council.

I ricercatori maschi pubblicano più saggi (*curva in nero*) delle donne (*in colore*) e la differenza aumenta con l'andare del tempo. Dati relativi a campioni estesi di persone che hanno conseguito il dottorato di ricerca fanno pensare che dagli anni quaranta la differenza si sia attenuata.

versità accreditate, situate nelle stesse zone geografiche degli scienziati eminenti: la scelta è stata effettuata avendo cura che gli elementi presi in considerazione avessero anche la stessa anzianità di servizio e svolgessero la loro attività nello stesso settore scientifico dei colleghi eminenti. Sebbene questi scienziati siano stati scelti in maniera sistematica, i criteri da noi adottati, il numero ridotto dei soggetti e l'esclusione da parte nostra di alcuni gruppi importanti (come, per esempio, i ricercatori occupati nell'industria) stanno a indicare che il nostro decisamente non è un vero e proprio campione casuale di tutti gli scienziati degli Stati Uniti.

Ai soggetti venivano poste domande sulla storia relativa al lavoro di ricerca e alle pubblicazioni e veniva chiesto di commentare alcuni grafici da noi preparati che mostravano tanto il numero di saggi che avevano pubblicato anno per anno, quanto importanti avvenimenti della loro carriera e della loro vita personale.

Gli scienziati, sia uomini sia donne, hanno riferito di essersi scontrati con l'opinione secondo la quale il matrimonio e la maternità non possono conciliarsi con una carriera scientifica impegnativa. Non desta sorpresa il fatto che il gruppo degli scienziati più anziani si sia scontrato con maggiore frequenza con questa opinione. Prima della seconda guerra mondiale (e anche poco dopo) era ampiamente diffusa la convinzione che le priorità peculiari di una donna vedessero al primo posto il matrimonio e la maternità e al secondo la scienza, mentre si pensava che una valida ricerca scientifica richiedesse tutta l'attenzione dello scienziato. L'idea che le donne potessero essere simultaneamente mogli e madri tradizionali e scienziati produttivi sembrava manifestamente assurda.

Queste opinioni sono state espresse chiaramente da una zoologa, la quale ricordava che «se una donna aveva dei figli e un marito che lavorava, non era previsto, sul piano psicologico, supporre che il suo lavoro fosse qualche cosa di più di un elemento di secondaria importanza». A quel tempo queste opinioni erano condivise da molti scienziati, sia uomini sia donne, i quali erano convinti che, una volta sposate, in generale le donne non potessero più essere scienziati seri. Una insigne biologa, che oggi ha più di settant'anni, ha raccontato che il suo capo di laboratorio, una donna, era rimasta atterrita alla notizia che la sua protetta si sarebbe sposata: «Mi buttò fuori del laboratorio nel momento stesso in cui apprese che avevo intenzione di prendere marito: per lei quello era un tradimento nei confronti delle donne.» Secondo un chimico, maschio, ai suoi tempi il matrimonio significava che, come scienziati, le donne erano «finite» e un fisico, anch'egli di sesso maschile, ha detto: «Non appena le donne iniziavano una vita domestica, era la fine per tutte quante.»

Questo clima culturale faceva sì che le donne decise a intraprendere seriamente la carriera di ricerca spesso non si sposassero. Nelle parole di una biologa che oggi ha superato la settantina «il matrimonio era considerato una cosa sconsigliabile (per le donne scienziato). Alla scienza ci si dedica anima e corpo. Si entra in un mondo tutto particolare, non si indossano vestiti normali... non ci si dovrebbe sposare, non si dovrebbero avere figli.»

Naturalmente non tutte le donne di scienza accettavano questi punti di vista, e alcune si sposavano, avevano figli e continuavano a lavorare. Almeno fino alla fine della prima guerra mondiale, però, il numero delle donne che si dedicavano alla ricerca era esiguo e meno della metà di esse era sposata. Le donne di scienza con marito e figli erano pressoché assenti dal mondo scientifico americano e si riteneva che le poche che vi appartenevano violassero le regole familiari prevalenti.

Sebbene l'atteggiamento della società nei confronti delle mogli e delle madri sia cambiato in misura significativa, anche le donne scienziato più giovani affermano che ancora oggi per molta gente il matrimonio e la maternità sono incompatibili con la carriera scientifica. Una giovane ricercatrice in chimica ci ha detto: «Quando restai incinta, fui cancellata dal novero degli scienziati seri... da moltissime persone.» Quando agiscono secondo queste convinzioni, coloro i quali occupano posizioni di potere limitano gravemente le opportunità e la carriera delle donne sposate.

Per valutare l'impatto effettivo del matrimonio e della maternità sulle donne attive in campo scientifico avevamo bisogno di una risposta a quattro domande. Come gruppo, le donne sposate sono ricercatrici meno produttive delle donne nubili? Fra le donne sposate, quelle che hanno figli pubblicano meno saggi di quelle senza figli? C'è una flessione, dopo il parto, nel rendimento delle ricercatrici dal punto di vista dell'attività pubblicistica? Il numero dei figli influisce sul rendimento di una donna che svolge attività di ricerca?

La storia delle pubblicazioni e della carriera di alcune donne appartenenti alla fascia degli scienziati eminenti ha offerto i primi indizi del fatto che in genere il matrimonio e i figli non influiscono minimamente sulla produttività scientifica. In media queste donne sposate eminenti (e le donne eminenti hanno esattamente le stesse probabilità di sposarsi e di avere figli delle loro controparti definite gregari) pubblicano nel corso della loro carriera un po' di più - non di meno - delle donne eminenti nubili: una media di 3,0 saggi l'anno contro 2,2. Fra le donne sposate eminenti, quelle con figli pubblicano 2,9 saggi l'anno, e quelle senza figli ne pubblicano 3,3. Inoltre, nel periodo di tre anni che precede e che segue la nascita del primo figlio la produttività annua dell'attività pubblicistica di queste donne non diminusce, ma anzi aumenta da 1,5 a 2,7 saggi. Infine, il tasso di pubblicazione di queste donne non ha alcuna attinenza con il numero dei figli.

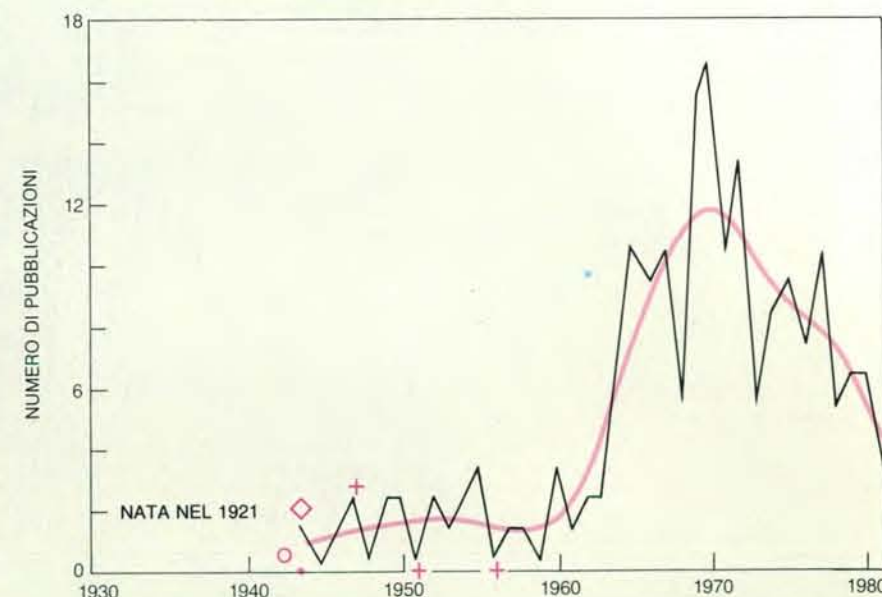
Queste risultanze statistiche sono palesemente controintuitive, ma sono in armonia con studi precedenti condotti su un campione rappresentativo dell'intera categoria. In quale misura corrispondono ai resoconti soggettivi fatti dalle donne nelle interviste? Le donne credono che sposarsi e avere figli non abbia davvero alcuna relazione con l'attività pubblicistica? E in tal caso, come spiegare questa convinzione?

La storia delle pubblicazioni di due donne scienziato eminenti più anziane, una con quattro e l'altra con tre figli, è riconducibile a uno schema generale (*si veda l'illustrazione in questa pagina*). In genere, queste donne attive in campo scientifico pubblicavano meno quando erano giovani e avevano figli piccoli; vi è un'accentuata tendenza ascendente nel numero di pubblicazioni edite dopo il primo decennio della loro carriera. All'interno di questa tendenza generale ascendente vi sono anche fluttuazioni - picchi e avvallamenti - anno per anno. Tutte queste donne riconoscono certamente che i figli richiedono molto tempo. Essi rappresentano un «preciso impegno in termini di tempo. Questo significa poter dedicare meno tempo ad altre attività», ma non si fa meno ricerca scientifica. La ricerca continua.

Come può continuare? In primo luogo, le donne eminenti pongono l'accento sulla circostanza che alla scienza si continua a pensare a casa così come al lavoro. Non si abbandona questo pensiero quando ci si chiude alle spalle la porta del laboratorio. «Quando i figli erano piccoli... mi venivano delle idee mentre lavavo i piatti e quando badavo ai bambini. Scientificamente parlando, ho svolto il lavoro migliore nel periodo in cui i bambini erano in arrivo.» In secondo luogo, se hanno un marito scienziato (e questa è la situazione tipica), esse parlano della ricerca durante «il cosiddetto tempo libero». In terzo luogo, gli obblighi professionali diversi dalla ricerca sono molto più limitati per le donne scienziato più giovani che per quelle più anziane. «Dedicavo più tempo alla scienza allora di quanto non faccia adesso... . Le richieste del mio tempo [allora] venivano dal lavoro e dai bambini. Ora vi sono molte altre cose.» Infine, il tasso di pubblicazione più ridotto che si riscontra nei primi anni non è necessariamente attribuibile alle esigenze della maternità, ma è caratteristico della fase iniziale di un programma di ricerca in via di sviluppo. Come ha osservato un fisico: «Nei primi anni stavamo costruendo quegli enormi strumenti. L'elaborazione della teoria e la sperimentazione [richiedevano] moltissimo tempo, e quindi non c'erano molti scritti da pubblicare.» Un altro, commentando le punte basse del grafico relativo alle sue pubblicazioni, ha detto: «Quando si scrive di meno vuol dire che si ha molto da fare.»

Secondo queste ricercatrici eminenti, il matrimonio e la maternità non riducevano la loro attività fatta poi oggetto di pubblicazione. Si deve credere a questi loro resoconti retrospettivi? Forse le loro percezioni non erano precise. Tutto considerato, l'esame dei grafici delle donne eminenti più anziane, sposate e con figli indica che in realtà esse erano scientificamente meno produttive quando erano giovani e avevano figli in tenera età. Il fatto di avere dei figli influiva realmente sul loro tasso di pubblicazione, almeno a breve scadenza?

A questo punto pensammo che l'esame della curva di produttività pubblicistica di due gruppi di scienziati sui quali il matrimonio e la condizione di genitore non dovrebbbero avere alcuna influenza - donne scienziato nubili e uomini eminenti le cui mogli si assumevano la responsabilità di accudire i figli - potessero illuminare ulteriormente queste prime risultanze controintuitive. Gli scienziati presumibilmente liberi da altri impegni pubblicano a un ritmo più rapido nei primi anni di attività rispetto alle donne che hanno figli piccoli? Probabilmente la risposta è negativa (*si veda l'illustrazione a pagina 76*). Nel primo decennio della loro carriera le donne nubili e gli uomini sposati hanno esattamente le stesse probabilità di produrre un numero ridotto di pubblicazioni. Ed è altrettanto probabile che con il passare del tempo essi presentino delle oscillazioni e un'inclinazione globale ascendente. Il fatto che le curve iniziali di produttività pubblicistica di questi due gruppi non differiscano molto da quelle delle donne sposate che hanno figli dà credibilità ai giudizi delle donne sposate.

Tuttavia, sorge un'altra domanda. Le donne eminenti più anziane asseriscono che il matrimonio e la gravidanza

Nei primi tre decenni della carriera le donne appartenenti alla fascia degli scienziati eminenti mostrano una generale tendenza ascendente nel numero di pubblicazioni annue. La produttività scientifica di due donne sposate, una con quattro (*in alto*) e l'altra con tre figli (*in basso*), segue lo stesso schema, non evidenziando effetti negativi durevoli del matrimonio o dei figli. La curva in nero indica il numero di scritti editi nei singoli anni, quella in colore la tendenza generale.

non avevano ridotto la loro produttività scientifica. Se le cose stanno in questo modo, per quale motivo si nota che il loro tasso di pubblicazione aumenta via via che esse superano l'età della cura dei figli, specie in considerazione delle distrazioni e delle responsabilità supplementari che a loro dire vengono con la maturità professionale e con un più alto grado di riconoscimenti? Parte della risposta va ricercata nel fatto che, via via che si avanza nella carriera, aumentano le opportunità di svolgere una ricerca in collaborazione con altri. Agli inizi i ricercatori eseguono da soli la maggior parte o tutto il lavoro sperimentale; quelli già affermati assumono spesso importanti ruoli amministrativi e sovrintendono il lavoro in corso nel loro laboratorio. Il curriculum relativo alle loro pubblicazioni rispecchia la conseguente crescita della ricerca svolta in collaborazione con altri.

E' bene menzionare il fatto che il curriculum delle pubblicazioni di alcune donne scienziato eminenti non presenta con l'andar del tempo la tipica inclinazione ascendente (*si veda l'illustrazione nella pagina a fronte*). Anche queste donne affermano che il matrimonio e la maternità non hanno influito in grande misura sulla loro produttività scientifica.

I nostri dati sembrano indicare quindi che in genere agli inizi della carriera le donne eminenti più anziane con marito e figli pubblicano quanto le loro controparti non sposate. E' tuttavia opportuno chiederci se sia possibile che da parte nostra sia stato commesso qualche errore cruciale nella compilazione e nell'interpretazione dei dati? Potrebbe darsi il caso che le donne che hanno figli ma che rimangono scientificamente produttive siano «autoselezionate», che siano cioè semplicemente scienziati di maggior talento rispetto alle donne che decidono di rimanere senza figli?

Sebbene non si possano fare paragoni rigorosi circa l'abilità scientifica, possiamo mettere a confronto i tassi di pubblicazione delle donne appartenenti alla fascia delle ricercatrici eminenti più anziane che rispettivamente avevano e non avevano figli, concentrandoci sugli anni precedenti la maternità. A questo scopo abbiamo accoppiato i due gruppi di donne grosso modo in base alle rispettive date di nascita. Il tasso di pubblicazione delle donne con figli relativo ai tre anni precedenti la nascita del loro primo bambino è stato messo a confronto con quello relativo ad «anni equivalenti» della vita delle donne senza figli. Abbiamo riscontrato storie iniziali analoghe: più o meno 1,3 saggi l'anno per le donne che successivamente hanno avuto figli e 1,6 per quelle che non ne hanno avuti. In altre parole, le donne eminenti più anziane che alla fine hanno avuto figli hanno pubblicato agli inizi della carriera un numero irrilevantemente minore di saggi rispetto alle donne che non hanno mai avuto figli.

E non potrebbe darsi, fatto ancora più importante, che sia stato un errore, da parte nostra, concentrare l'attenzione sulla storia di donne del gruppo degli scienziati eminenti anziché su quelle di donne che avrebbero potuto avere maggiori probabilità di sperimentare gli effetti negativi del matrimonio e della maternità sul tasso di pubblicazione? Tutto sommato, le donne eminenti sono scienziati di grande successo; se nel loro caso il matrimonio e la maternità avessero imposto un tributo, presumibilmente non sarebbero riuscite a ottenere quei riconoscimenti che in realtà hanno avuto. La storia delle pubblicazioni di altre donne attive in campo scientifico, quelle che rientrano nella categoria dei gregari, attesta forse l'influsso negativo della maternità e della cura dei figli?

Ovviamente gli scienziati, di entrambi i sessi, che abbiamo definito gregari pubblicano in generale meno dei loro colleghi eminenti. All'interno dei gregari del nostro campione, le donne sposate hanno effettivamente pubblicato saggi in misura di poco inferiore alle donne nubili (una media di 1,1 contro 1,7 l'anno).

Un tasso di pubblicazione più ridotto nella prima parte di una carriera è caratteristico sia degli uomini sposati sia delle donne nubili. Il profilo delle pubblicazioni di un'insigne biologa (*in alto*) che non si è mai sposata presenta lo stesso schema di oscillazioni e lo stesso aumento globale rilevabili nei grafici relativi alle donne che si sono sposate e che hanno avuto figli. Lo stesso schema si riscontra nel profilo di un chimico eminente (*in basso*), il quale ha pubblicato a un ritmo molto più lento quando i figli erano piccoli, sebbene le sue responsabilità domestiche fossero minime.

Ma le donne sposate e con figli non pubblicano meno saggi di quelle sposate senza figli; la media di entrambi i gruppi è di circa un saggio l'anno. Come nel caso delle donne eminenti, il tasso di pubblicazione dei gregari non ha esibito alcuna flessione dopo la nascita dei figli. Le donne di quest'ultimo gruppo hanno pubblicato molto meno di un saggio l'anno (0,2) nei tre anni precedenti la nascita del primo figlio e poco meno di un saggio (0,8) l'anno nei tre anni successivi.

Dalle dichiarazioni di queste donne scienziato si ricava più o meno la stessa impressione: il fatto di avere figli non ha inciso in misura significativa sul loro curriculum relativo alla ricerca e alle pubblicazioni. Come ha fatto notare una studiosa di scienza del comportamento relativamente improduttiva: «Non mi è passato neanche per la testa di smettere di lavorare quando ho avuto [una figlia]... . Anzi, subito dopo la nascita della bambina ho scritto un saggio e ho iniziato a lavorare a quello successivo... quindi, se mai... ho avuto l'impressione di lavorare meglio (che è come dire più efficientemente anche se con un carico superiore).»

Queste parole sono in armonia con quelle di altre donne, come una biochimica secondo la quale il suo tasso di pubblicazione non era influenzato dagli impegni familiari. «È solo fortuito... tra i miei ruoli di madre e di ricercatrice non esiste un rapporto diretto... . Una cosa non ha interferito con l'altra.» Contrariamente alle aspettative, quindi, non è probabile sentir dire da queste donne, più di quanto non lo sia per quelle eminenti più anziane, che il matrimonio e le responsabilità familiari influiscono sul tasso di pubblicazione, e i dati statistici in nostro possesso confortano queste affermazioni.

Alcuni scienziati pubblicano a un ritmo abbastanza costante lungo tutto l'arco della carriera. Questi sono i profili di due donne eminenti: una che si è sposata due volte e ha avuto due figli (*in alto*) e un'altra che non si è mai sposata (*in basso*). Il numero annuo delle pubblicazioni oscilla, ma la media misurata per periodi di cinque anni rimane per queste donne più o meno la stessa. Questo schema appare con la stessa frequenza fra le donne sposate e quelle non sposate.

Troveremmo uno schema analogo nei dati relativi alle donne più giovani? Dato che di solito il matrimonio e la gravidanza vengono agli inizi della carriera, i dati relativi alle donne più giovani dovrebbero dimostrare i loro effetti sulla produttività in fatto di pubblicazioni, almeno a breve scadenza.

Una studiosa di scienza del comportamento, oggi professore ordinario in un dipartimento di prim'ordine e che aveva appena avuto un bambino, dichiarò che la maternità non aveva il minimo rapporto con l'andamento delle sue pubblicazioni. «Avere un figlio è logorante sotto molti aspetti, ma non ha inciso certo sul mio lavoro, specie se considero quanto ho... fatto quest'anno. No, davvero, sono il cinema, la vita sociale e cose del genere che contano da questo punto di vista... . Mi sento cronicamente inoperosa e in arretrato, e quest'anno ne sto dando la colpa al bambino, ma mi rendo conto che la cosa non ha niente a che fare con lui.» In passato era «il numero eccessivo di studenti... [un] comitato di revisione delle borse di studio, [un] lavoro editoriale e [adesso] un bambino. Così ho sempre avuto una sorta di bambino su cui scaricare ogni colpa.»

Forse il caso limite è quello della titolare di una cattedra in un dipartimento importante. Questa donna si è sposata quattro volte, ha divorziato tre volte e ha avuto quattro figli da tre mariti diversi. Se il matrimonio e la maternità dovessero davvero interrompere la carriera scientifica, in questo caso lo avrebbero fatto certamente. La produzione pubblicistica di questa donna è andata invece sempre aumentando lungo tutto l'arco della sua complessa storia personale. Ironia della sorte, la maggiore flessione nella sua attività si ebbe nel 1979, uno dei pochi anni in cui non si sposò, non ebbe figli e non divorziò (*si veda l'illustrazione nella pagina successiva*). A una domanda sull'andamento delle sue pubblicazioni, la donna rispose: «Tutt'a un tratto sei pronta a riferire su tre progetti diversi e quindi [i saggi] escono dalle rotative... . Gli alti e bassi [non hanno] niente a che fare con il resto della mia vita.»

Nondimeno, la gravidanza e i suoi postumi hanno temporaneamente interferito con la ricerca nel caso di tre delle 37 donne con figli del nostro campione. Una biologa ci disse: «Io ero una di quelle donne che dicevano di essere ben organizzate, che non avrebbero fatto altro che mettere al mondo un bambino e basta - ma non mi rendevo conto che gli ormoni potessero rendere un simile servizio a una persona.» Alla richiesta di spiegare la ragione per la quale il suo tasso di pubblicazione aveva registrato una flessione solo temporanea e non molto accentuata, rispose: «Sono stata fortunata perché a quel tempo avevo delle persone nel mio laboratorio. Queste persone erano attive [e] produttive... . Ma mi resi conto che per un anno intero la mia testa non funzionava.»

Questi dati «longitudinali» indicano che, in complesso, matrimonio e figli non sono in contrasto con l'attività pubblicistica delle donne. Anche se poche donne scienziato metterebbero in discussione il fatto che il matrimonio e la maternità impongono terribili responsa-

bilità, a quanto pare molte riescono a cavarsela simultaneamente con una carriera e con gli impegni familiari. Come fanno? E quando sorgono dei conflitti fra casa e carriera, come li affrontano?

La risposta va ricercata in parte nel modo in cui le donne che fanno ricerca gestiscono l'insieme di ruoli sociali che ognuna di esse riveste (come docente, direttore di laboratorio, moglie e cittadina). Noi ci siamo concentrati su tre aspetti di questo complesso di ruoli collegati fra loro: totale complessivo (il numero delle posizioni occupate simultaneamente), congruenza (la misura in cui vari obblighi sono in armonia e non in conflitto fra di loro) e la scelta del momento più opportuno per aggiungere o eliminare alcuni impegni legati alla posizione sociale.

Come caso estremo, parecchie donne, convinte che il matrimonio sia incompatibile con l'attività scientifica, decidevano di non sposarsi, limitando così la gamma di ruoli sociali occupabili. Tre quarti delle donne del nostro campione si sono sposate, una percentuale che oggi in genere sembra tipica per le donne che operano in questo settore. Per la maggior parte di esse era fondamentale, a proposito del matrimonio, la scelta del momento più opportuno. Come disse una giovane economista: «Per me [sposarmi] sarebbe anche un grosso svantaggio dal punto di vista della carriera. Una volta che diventassi di ruolo e fossi sistemata in un'università, sarebbe un po' più facile.»

Due terzi di queste donne sposate avevano figli. Scegliere il momento più opportuno per il loro arrivo, dicono molto donne, aiuta a conservare una carriera nel campo della ricerca. Una famosa studiosa di fisica rimandò per nove anni dal matrimonio l'arrivo del primo figlio in modo da potersi mettere alla prova come professionista. Molte donne più giovani dichiararono che stavano rimandando la maternità fino al momento in cui fossero diventate di ruolo. Nel dubbio di poter avere un bambino e nel contempo mantenere il livello di rendimento necessario per ottenere il ruolo, una giovane biochimica disse: «Il mio scenario ideale è quello di ottenere un posto di ruolo e successivamente avere uno o due figli.»

In complesso, donne eminenti e donne gregarie avevano più o meno lo stesso numero di figli: due in media, con nessuna che superava i quattro. I dati da noi ottenuti indicano che il tasso annuale di pubblicazione è sostanzialmente lo stesso per le donne che hanno un unico figlio e per quelle che ne hanno due o più.

Vi sono aspetti del matrimonio e della maternità diversi dalla scelta del momento più opportuno che possono favorire un insieme congruente di ruoli sociali. Quasi quattro quinti delle donne da noi intervistate erano sposati con uno scienziato (anche in questo caso la proporzione è tipica delle donne di scienza in genere). A quanto pare, questo accoppiamento non casuale, o «selettivo», dà a queste donne (e anche agli uomini) una molteplicità di benefici, inclusa la sollecita comprensione per gli impegni professionali e il modo di vivere. Una studiosa di biologia molecolare osservò che il marito non si sarebbe certo preoccupato se lei tornava a casa tardi perché «sapeva che, per quanto io avessi pianificato bene qualche cosa, gli esperimenti subiscono sempre qualche ritardo. Penso che questo abbia facilitato molto le cose.» Le donne ricercatrici sposate con uno scienziato pubblicano, in media, il 40 per cento in più delle donne sposate con un uomo che ha un'altra occupazione. Non è da escludere che questa differenza nel tasso di pubblicazione sia dovuta ad autoselezione, a congruenza di valori o alla flessibilità dei programmi accademici.

Le donne impegnate nella ricerca possono raggiungere la congruenza fra gli obblighi dei diversi ruoli dividendo la loro vita in compartimenti stagni, ma a loro dire questa divisione non è sempre possibile. Anzi, molte trovano più difficile non pensare al lavoro quando sono a casa di quanto lo sia non pensare ai figli quando sono al lavoro. Tutte le donne con figli mettevano in evidenza il fatto che dovevano contare su qualche forma di aiuto per la cura dei figli o per la casa, soluzioni necessarie ma, nel migliore dei casi, provvisorie. La malattia di uno dei due coniugi, di un figlio o della governante può mandare a monte tutti i piani.

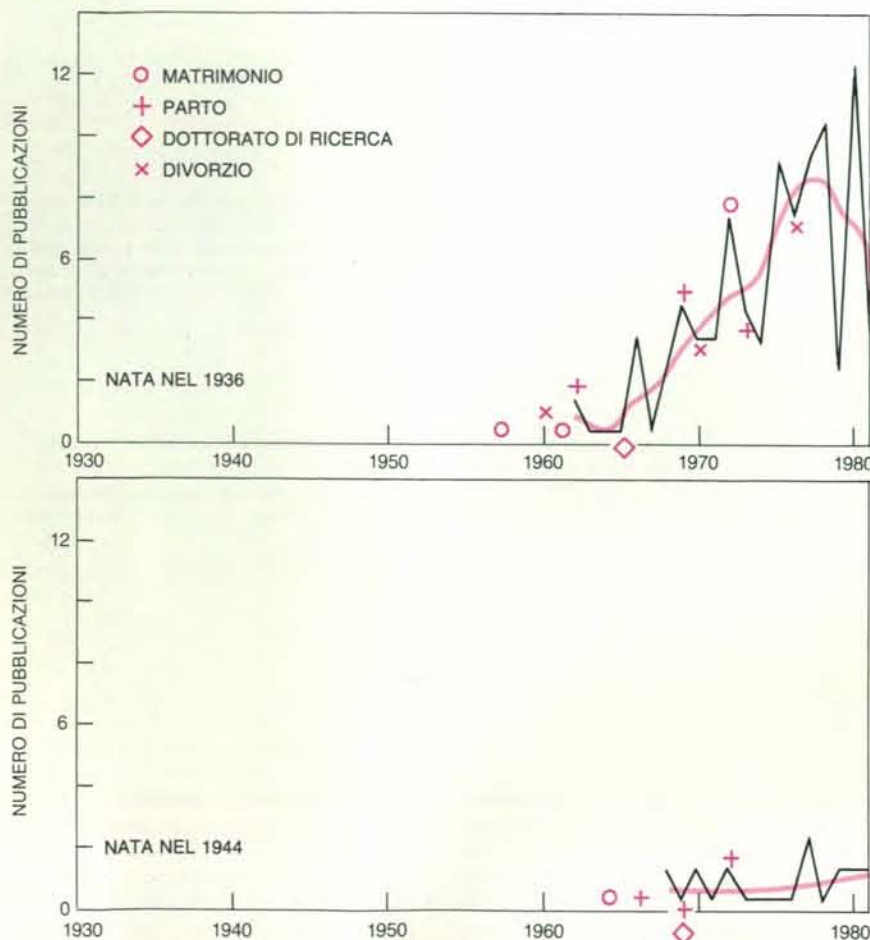
Considerate queste difficoltà, è possibile che gli scienziati di entrambi i sessi riescano a continuare il lavoro di ricerca solo trascurando consorte e figli? Il nostro studio non era inteso a rispondere a questa domanda. Sappiamo però che il tasso di divorzi tanto per le donne quanto per gli uomini non ha alcun rapporto con la produttività pubblicistica.

Le donne ricercatrici con marito e figli pagano effettivamente un prezzo per rimanere scientificamente produttive. Esse dichiarano di aver dovuto eliminare quasi tutto al di fuori del lavoro e della famiglia, specie quando i figli erano piccoli. Come ha messo in rilievo un'eminente psicologa, a essere sacrificato per primo è «il tempo libero. Penso di poter lavorare efficacemente soltanto... 50 ore la settimana... . Se non avessi figli, probabilmente leggerei più romanzi... o andrei più spesso al cinema.»

La mancanza di tempo libero non solo influisce sugli svaghi, ma a volte ha anche gravi conseguenze per la ricerca e per la carriera delle donne, anche se non ha alcun effetto significativo sul tasso di pubblicazione. Le donne che si mantengono fedeli a rigorosi programmi familiari sostengono di aver perduto la flessibilità che le portava a trattenersi in laboratorio fino a tarda ora per lavorare a un problema interessante; esse affermano di non sentirsi parte del «club», di non avere tempo per discussioni informali con i colleghi.

Altri ricercatori hanno dimostrato che solo il 12 per cento delle donne che svolge attività scientifica smette di lavorare dopo aver conseguito il dottorato di ricerca. Certamente, per alcune, questo accade a causa degli intensi conflitti che insorgono tra attività scientifica e maternità. Una donna che aveva rinunciato a una promettente carriera di ricercatrice per un lavoro amministrativo disse in un'intervista supplementare: «Frequentavo il laboratorio solo nelle ore in cui i bambini erano a scuola... . Lavoravo con persone davvero molto intelligenti che sfornavano pubblicazioni a un ritmo che non riuscivo a sostenere... . La cosa era troppo deludente.» Un piccolo sottogruppo di donne ritiene che scienza e maternità non siano conciliabili e modifica la propria carriera in modo da poter dedicare più tempo alla famiglia.

Il nostro studio indica che, per la maggior parte di queste donne, scienza e maternità possono coesistere. Le donne scienziate che si sposano e hanno famiglia, in media danno alle stampe ogni anno lo stesso numero di pubblicazioni delle donne nubili. Riuscire a soddisfare le esigenze simultanee della carriera di ricerca, del matrimonio e della maternità non è facile; è qualche cosa che richiede organizzazione e un elaborato insieme di adattamenti personali.

I risultati di questo studio non dovrebbero essere interpretati nel senso che matrimonio e figli non hanno alcun effetto sulla carriera delle donne attive in campo scientifico. Un effetto ce l'hanno, ma in genere non impongono un tributo al rendimento delle donne come ricercatori. Come si può spiegare allora la persistente disparità del tasso di pubblicazione tra scienziati uomini e donne? Perché nel corso della loro carriera gli uomini pubblicano più saggi delle donne che hanno una preparazione analoga? La differenza non si spiega semplicemente con il matrimonio e la maternità, ma rimane un enigma che richiede ulteriori indagini.

La produttività pubblicistica non ha il minimo rapporto con gli impegni familiari, come risulta dalla storia delle pubblicazioni di due donne la cui carriera presenta un andamento notevolmente diverso. Una (*in alto*) è un'eminente studiosa di scienza del comportamento la cui produttività è aumentata costantemente nonostante quattro matrimoni (l'ultimo nel 1982), tre divorzi e la nascita di quattro figli. L'altra (*in basso*) è professore associato di chimica e ha tre figli. Quest'ultima attribuisce la propria bassa produttività a vincoli diversi dagli impegni domestici. La storia delle sue pubblicazioni è tipica di molte ricercatrici da noi definite «gregarie».

Il calcolatore come strumento musicale

Generando l'opportuna successione di numeri, un calcolatore può produrre qualunque suono, anche mai udito prima; grazie a questa versatilità, la sintesi digitale si è conquistata un posto nella musica

di Max V. Mathews e John R. Pierce

Lo scorso anno il compositore e direttore d'orchestra francese Pierre Boulez ha sorpreso gli appassionati americani compiendo una tournée per il paese con una formazione che comprendeva un calcolatore e una schiera di dispositivi elettronici per l'elaborazione dei suoni. Le opere di Boulez per orchestra e suoni generati dal calcolatore integrano apparecchiature elettroniche digitali in un ruolo attivo, di produzione sonora, che trascende di gran lunga il ruolo più convenzionale attribuito a queste apparecchiature nella registrazione e nella riproduzione di musica sinfonica. I calcolatori non sono ancora membri stabili di ogni orchestra sinfonica, ma già i sintetizzatori digitali fungono da alternativa agli strumenti musicali tradizionali nella produzione di colonne sonore per il cinema e la televisione e si avviano a diventare con grande rapidità gli strumenti d'elezione della musica «leggera».

A monte di tutti i tentativi di inserire nel mondo della musica la sintesi digitale di suoni complessi, stanno alcune ricerche pionieristiche sull'elaborazione dei suoni al calcolatore, in cui ambedue gli autori hanno avuto il privilegio di essere coinvolti, e che hanno preso il via negli anni cinquanta presso i Bell Telephone Laboratories. A tutta prima eravamo stati attirati verso il calcolatore come dispositivo per l'analisi e la produzione di suoni, mentre studiavamo i fattori che contribuiscono alla trasmissione efficace del parlato lungo le linee telefoniche. Ci fu presto chiaro che la qualità dei suoni è di grande importanza non solo per il linguaggio parlato, ma anche per la musica, e con grande entusiasmo abbiamo cominciato a dedicarci allo studio della produzione dei suoni musicali.

I nostri primi tentativi di produrre suoni musicali con un calcolatore furono deludenti, ma gli strumenti elettronici e i programmi di calcolatore che si sono evoluti da questi tentativi iniziali oggi sono abbastanza raffinati da poter avere un'influenza significativa in molti settori attinenti alla musica. In primo luogo, gli strumenti sono importanti, sotto il profilo commerciale, per l'industria della registrazione sonora, perché possono generare efficacemente musica o effetti sonori di facile sincronizzazione con l'azione rappresentata in una pellicola cinematografica o in televisione. In secondo luogo, mettono a disposizione un universo virtualmente illimitato di suoni, mediante i quali compositori ed esecutori possono esprimere pensieri e sentimenti. Infine, ed è questa forse la cosa più importante, possono approfondire la nostra comprensione di quelle particolari configurazioni sonore che chiamiamo «musica».

Ridotto alla sua natura fisica essenziale, un suono altro non è che una fluttuazione di pressione nell'aria. Come tale, può essere rappresentato graficamente per mezzo di una forma d'onda: un grafico che dà la variazione della pressione in funzione del tempo. Il modo in cui il suono viene percepito (il fatto cioè che sia o meno piacevole) dipende dal modo in cui le varie caratteristiche delle fluttuazioni di pressione vengono tradotte in impulsi nervosi nell'orecchio e dal modo in cui gli impulsi nervosi vengono poi interpretati soggettivamente dal cervello.

I suoni che vengono percepiti come dotati di un'altezza ben definita hanno forme d'onda che presentano una variazione quasi periodica della pressione. L'altezza di un suono corrisponde direttamente alla frequenza con cui la variazione si ripete. Per esempio, una variazione di pressione che si ripeta 440 volte al secondo viene percepita da un musicista come un suono con un'altezza ben definita: un la subito sopra il do centrale.

La maggior parte dei suoni che udiamo ha origine da vibrazioni indotte in oggetti comuni: le corde vocali umane attraverso le quali viene emessa l'aria, le corde di violino sfregate da un archetto o alcune automobili che si scontrano, tanto per citare qualche caso. I suoni possono essere generati anche dalle vibrazioni indotte in un altoparlante facendo variare la tensione ai suoi morsetti d'ingresso. Infatti, come sa ogni audiofilo, si può ottenere una eccellente riproduzione di un dato suono applicando a un buon altoparlante una tensione che riproduca fedelmente la funzione di pressione acustica.

Un enunciato matematico, che va sotto il nome di teorema del campionamento, dice che qualunque forma d'onda costituita da più componenti di frequenze diverse può essere descritta esattamente da una successione di numeri che danno il valore dell'ampiezza della forma d'onda a intervalli determinati dalla larghezza di banda della forma d'onda stessa, cioè dalla gamma delle frequenze componenti (convenzionalmente espresse in hertz, ovvero cicli al secondo). Specificamente, la frequenza di campionamento (la velocità con cui debbono essere generati i numeri) deve essere uguale al doppio della larghezza di banda della forma d'onda. Questo teorema, dimostrato da Claude E. Shannon ai Bell Laboratories nel 1948, è alla base di tutte le tecniche digitali di registrazione, elaborazione e generazione dei suoni.

Il teorema del campionamento comporta che si possa registrare esattamente un secondo di un segnale acustico con una larghezza di banda di 20 000 hertz (che copre tutta la gamma delle frequenze udibili dall'uomo) se, nell'arco di quel secondo, si raccolgono 40 000 numeri (chiamati campioni) che corrispondono ai valori istantanei, a intervalli regolari, dell'ampiezza della pressione acustica (o della tensione elettrica a essa corrispondente). Di converso, se si possono richiamare gli opportuni 40 000 valori campionati al secondo, si potrebbe riprodurre qualunque suono percepibile, in tutta la sua complessità acustica. Il compact disc, in cui vengono codificati e immagazzinati, sotto forma di punti di riflettenza variabile, circa 40 000 campioni di suono al secondo, costituisce un esempio di un simile sistema di archiviazione e recupero.

Un altro metodo per archiviare e recuperare queste grandi quantità di numeri ai fini della generazione di suoni ci è offerto dai microelaboratori digitali, come quelli che si trovano nei calcolatori. In un calcolatore la conversione dei numeri in tensioni - un passo essenziale nella produzione digitale di suoni - può essere effettuata facilmente mediante convertitori da analogico a digitale, che traducono un segnale elettrico in una successione di numeri proporzionali alla tensione del segnale, e mediante convertitori da digitale ad analogico, che svolgono il processo opposto. La successione di impulsi discreti di tensione che un convertitore da digitale ad analogico produce da una successione di campioni di ampiezza, in genere viene «arrotondata» in una forma d'onda continua mediante un filtro speciale, prima che il segnale elettrico sia amplificato e riprodotto attraverso un sistema di altoparlanti.

Per motivi pratici, sia i calcolatori, sia i compact disc esprimono i campioni di ampiezza in termini di numeri binari. Lo standard attuale prevede la definizione di un campione per mezzo di 16 bit (16 cifre binarie). In questo modo l'ampiezza di una forma d'onda può essere suddivisa in 65 536 livelli discreti. (Nella pratica, metà di questi numeri binari sono utilizzati per rappresentare campioni di ampiezza positiva e l'altra metà per campioni di ampiezza negativa.) Questa gamma di livelli d'ampiezza non è sufficiente per una buona riproduzione di concerti di musica rock o di musica sinfonica che sia totalmente priva di rumore percettibile, ma la fedeltà è sicuramente molto superiore a quella delle registrazioni analogiche sui normali dischi fonografici o su nastro magnetico.

I nostri tentativi iniziali di applicare il teorema del campionamento per generare suoni mediante un calcolatore ci diedero una sorpresa poco piacevole: sembrava che fossimo in grado di sintetizzare solo suoni non musicali, anziché suoni musicali piacevoli. Quei primi tentativi consistevano soprattutto nella trasformazione in suoni di configurazioni numeriche che rappresentavano forme d'onda semplici (come la sinusoide o il dente di sega), e i suoni che riuscivamo a ottenere erano per lo più o scialbi o ronzanti ed «elettronici».

Non che il calcolatore con cui lavoravamo fosse di per sé limitato nelle forme d'onda che poteva produrre (anche se avevamo qualche vincolo pratico, in termini di denaro e di tempo); il problema piuttosto era che nessuno sapeva quali potessero essere le forme d'onda che l'o-

L'elaboratore digitale dei suoni fa ormai parte integrante e attiva dell'organico dell'Ensemble InterContemporain, qui fotografato durante le prove di *Répons*, una composizione del musicista francese Pierre Boulez. I tecnici (*in primo piano*) controllano l'elaborazione in tempo reale (ovvero istantanea) del suono da una stazione di lavoro elettronica, guidati da Boulez stesso (*sullo sfondo*), che dirige l'esecuzione.

recchio umano avrebbe percepito come suoni musicali piacevoli. Una qualità caratteristica del suono è il suo timbro, ovvero il suo «colore». Riuscivamo a generare con il calcolatore suoni di altezza e intensità date, ma si dimostrò difficile generare suoni di timbro piacevole.

La letteratura sull'acustica degli strumenti tradizionali era molto scarsa, e non ci fu di grande aiuto: si rivelò non solo incompleta, ma in molti casi addirittura clamorosamente errata. Per esempio, anche se nella letteratura veniva messa particolarmente in rilievo la descrizione di un suono musicale in termini del suo stato stazionario, cioè della sezione centrale della sua forma d'onda (la fase di regime del suono), ci risultò presto chiaro che erano più importanti l'inizio della forma d'onda (l'attacco) e la sua fine (l'estinzione). Il contorno di una forma d'onda, tracciato collegando cresta a cresta e ventre a ventre, è detto il suo inviluppo. Un inviluppo che abbia un attacco netto, in rapida ascesa, seguito da una fase di estinzione di andamento morbido, dà il timbro di un suono pizzicato oppure ottenuto per percussione, indipendentemente dalle fluttuazioni di stato stazionario della forma d'onda all'interno del suo inviluppo.

Si pensava inoltre che la conoscenza delle ampiezze relative delle varie componenti di frequenza dello stato stazionario di un suono musicale fosse sufficiente per caratterizzarne il timbro. Uno spettro di frequenza di un suono musicale mostra che la sua forma d'onda contiene non soltanto una componente alla frequenza corrispondente all'altezza (quella che in genere viene definita frequenza fondamentale), ma numerose componenti le cui frequenze in genere, anche se non sempre, sono multipli interi della fondamentale. Queste componenti sono chiamate frequenze parziali (o, per brevità, semplicemente parziali) e possono essere ordinate in base alla loro frequenza. La frequenza fondamentale è quindi la prima parziale, la componente di frequenza immediatamente superiore è la seconda parziale, e via di questo passo.

Secondo la concezione tradizionale del timbro, per esempio, il suono di un clarinetto sarebbe completamente determinato da uno spettro in cui le frequenze di tutte le parziali fossero multipli dispari della fondamentale. Questa proprietà dello spettro di frequenza di un clarinetto, in una certa misura, è vera, ma abbiamo trovato che non è sufficiente a caratterizzare il timbro dello strumento.

Sono state necessarie molte ricerche particolareggiate per capire gli aspetti timbrici del suono e per correggere alcune concezioni antiquate e fuorvianti. Per nostra fortuna il calcolatore stesso si è dimostrato un ausilio potente per lo studio dei timbri musicali. Una notevole dimostrazione di come si debba procedere per analizzare e sintetizzare un suono con un buon timbro è offerta dal primo lavoro svolto nel 1965 ai Bell Laboratories dal compositore e fisico francese Jean-Claude Risset.

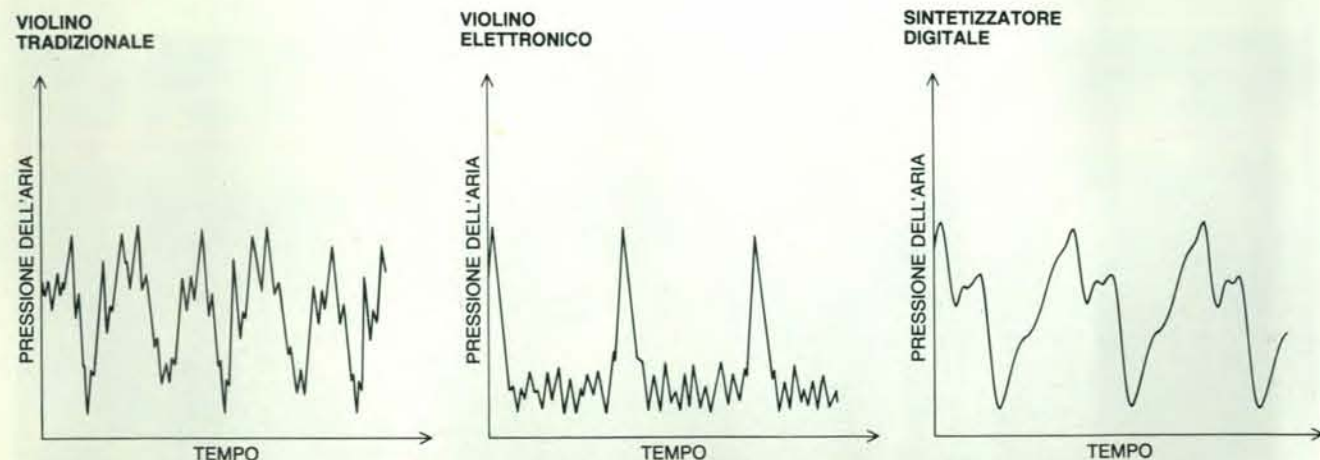
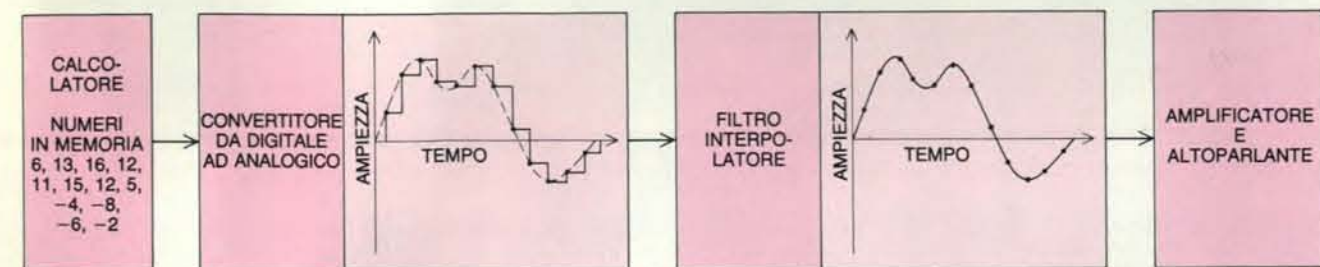
Risset cercava di sintetizzare buoni suoni di ottoni basandosi su quel che aveva appreso nella letteratura in merito alle ampiezze relative delle parziali di una tromba. Programmò un calcolatore in modo che generasse numeri corrispondenti ai campioni di ampiezza di quella che pensava sarebbe stata una forma d'onda da tromba. Quando però trasformò i numeri in suoni, trovò che il risultato non assomigliava affatto al suono dello strumento tradizionale.

Risset allora registrò i suoni di una tromba reale e ne analizzò gli spettri con un calcolatore. Scoprì così che lo spettro di una tromba cambia durante l'esecuzione di un suono, e che le parziali di frequenza più alta hanno ampiezza maggiore nella fase centrale che non in quelle di attacco e di estinzione. Sintetizzando un suono le cui parziali di frequenza più alta crescevano lentamente nella fase di attacco e raggiungevano la loro ampiezza massima durante lo stato stazionario, riuscì a produrre suoni che un ascoltatore medio non avrebbe potuto distinguere dai suoni registrati di una tromba.

Per farla breve, Risset aveva scoperto che lo spettro di frequenza nella fase di stato stazionario non costituisce una descrizione adeguata di un timbro. Per sintetizzare un suono musicale di un particolare timbro si deve sapere come varia lo spettro di frequenza durante l'esecuzione di una nota, cioè come crescono le varie parziali all'inizio e come si estinguono alla fine. Inoltre Risset ha scoperto che parziali diverse seguono un andamento diverso nel tempo, e che questo è di importanza critica per l'ascoltatore. Pertanto uno spettro di frequenza di stato stazionario a cui sia data forma mediante un unico inviluppo ben di rado è sufficiente per riprodurre il suono di strumenti tradizionali.

Il lavoro di Risset ha seguito una metodologia denominata analisi per sintesi, che ha portato a grandi progressi nella comprensione del timbro dei suoni prodotti da strumenti tradizionali. La metodologia dell'analisi per sintesi parte con un'analisi di un suono, spesso effettuata con un calcolatore che può suddividere il suono nelle singole componenti di frequenza e determinare l'inviluppo di queste. Quest'analisi di solito produce una enorme quantità di dettagli, da cui lo sperimentatore deve scegliere le caratteristiche che ritiene importanti per la produzione del timbro tipico di un suono. Il ricercatore, quindi, formula un'ipotesi che dà una semplice descrizione fisica del suono e la mette alla prova sintetizzando un suono su questa base. L'ipotesi viene valutata mettendo a confronto il suono di sintesi con il suono originale: se sono indistinguibili per l'ascoltatore, allora lo sperimentatore ha avuto successo nel selezionare quei dettagli che caratterizzano effettivamente una particolare classe di suoni.

Per sintetizzare suoni, poco importa se a scopo di analisi o a scopo musicale, bisogna avere a disposizione software che generi efficacemente la successione dei numeri binari che rappresentano i campioni successivi di una forma d'onda. Il software deve essere di semplice impiego e al tempo stesso abbastanza raffinato da consentire di dar corpo a quasi ogni ipotesi sul timbro di un suono mediante una forma d'onda.

Fra la fine degli anni cinquanta e i primi anni sessanta, uno di noi (Mathews) ha scritto un software di questo tipo. Uno dei programmi costitutivi, Music V, contiene i concetti fondamentali che si trovano tuttora nel software utilizzato dai centri di musica al calcolatore in tutto il mondo. Ne fanno parte numerosi semplici blocchi costruttivi di programma, chiamati generatori, che danno al musi-

Le forme d'onda indicano come un particolare suono provochi la fluttuazione della pressione dell'aria ambiente. Il suono prodotto da un violino tradizionale (*a sinistra*) è caratterizzato da una forma d'onda periodica complessa. Un violino elettronico (*al centro*) costruito da uno degli autori (Mathews) converte il movimento di una corda metallica sfregata dall'archetto in un segnale elettrico; quest'ultimo viene quindi filtrato in modo da ottenere una semplice forma d'onda analoga a quella del violino. Con un sintetizzatore digitale di larga diffusione (*a destra*) è possibile imitare la forma d'onda di un suono reale di violino più strettamente di quel che possa fare il violino elettronico.

Il teorema del campionamento afferma che qualunque forma d'onda costituita da componenti di frequenze diverse può essere descritta esattamente da una successione di numeri. Questo teorema sta alla base di tutte le tecniche di elaborazione digitale dei suoni. I numeri, chiamati campioni, sono proporzionali all'ampiezza istantanea della forma d'onda e la frequenza di campionamento minima è pari al doppio della larghezza di banda (la gamma delle frequenze componenti) della forma d'onda. I campioni archiviati nella memoria di un calcolatore possono essere trasformati in tensioni, proporzionali al valore di ciascun campione. Le tensioni discrete poi possono essere interpolate in modo da dare un segnale continuo che, opportunamente amplificato, viene inviato a un altoparlante e convertito nuovamente in energia sonora. La sintesi dei suoni comporta la programmazione di un calcolatore, affinché produca l'opportuna successione di numeri per un suono dato.

ISTANTE DI INIZIO (SECONDI)	DURATA (SECONDI)	INTENSITÀ (UNITÀ ARBITRARIE)	FREQUENZA (HERTZ)
0,0	1,5	1	262
2,0	0,25	5	325

Uno «strumento» al calcolatore è costruito partendo da unità di generazione nel programma di sintesi dei suoni Music V, scritto da uno degli autori (Mathews). Le unità di generazione sono sottoprogrammi i cui ingressi e le cui uscite numeriche possono essere variamente collegati. L'unità di generazione più importante è l'oscillatore. Ogni volta che un oscillatore compie un ciclo, genera una serie di numeri che corrisponde a una forma d'onda scelta. L'ampiezza della forma d'onda in uscita e la frequenza dei cicli di generazione della forma d'onda sono determinate dai due ingressi dell'oscillatore. L'ingresso relativo all'ampiezza di un oscillatore che produce suoni di altezza determinata spesso è l'uscita di un altro oscillatore, che controlla l'inviluppo del suono. L'inviluppo determina l'attacco del suono (cioè la velocità con cui si raggiunge l'ampiezza di regime), il suo stato stazionario (la parte centrale, o fase di regime) e la sua estinzione (il modo e la velocità con cui cessa o «decade»). Uno strumento così costruito viene «suonato» mediante liste di note (*in basso a sinistra*): istruzioni per il calcolatore che specificano sostanzialmente le stesse informazioni che una nota, sul rigo musicale (*in basso a destra*), indica all'esecutore.

cista la flessibilità per creare suoni strumentali unici, tabelle numeriche memorizzate per generare efficacemente alcune forme d'onda a comando e liste di note (al posto della notazione musicale convenzionale) che specificano quello che deve essere eseguito.

Un compositore può costruire i suoi «strumenti» musicali collegando le unità di generazione in molti modi diversi. Per certi versi le unità di generazione sono simulazioni al calcolatore dei dispositivi analogici presenti nei primi sintetizzatori elettronici, che l'utente collegava fra loro con opportuni cavetti. Mentre i dispositivi modulari di un sintetizzatore analogico manipolano tensioni elettriche, le unità di generazione di Music V manipolano numeri. Le unità di generazione più importanti sono l'oscillatore, il sommatore e il moltiplicatore.

Ogni volta che viene attivato, un oscillatore genera una serie di numeri che corrisponde a una forma d'onda tratta

Due metodi per generare forme d'onda musicali complesse sono la sintesi additiva e quella in modulazione di frequenza (FM, dall'inglese *Frequency Modulation*). Nella prima le forme d'onda vengono sommate per produrre una forma d'onda musicale. Le frequenze delle forme d'onda sommate in genere sono multipli interi della frequenza che determina l'altezza e corrispondono alle componenti di frequenza di un suono, chiamate parziali. Nella sintesi in FM, la frequenza di una forma d'onda (la portante) viene modulata dall'ampiezza di un'altra forma d'onda (la modulante). Le forme d'onda risultanti sono più complesse di quelle di partenza e producono timbri più ricchi. Timbri ancora più complessi si ottengono variando le ampiezze relative e le fasi di portante e modulante durante l'esecuzione di una nota. La sintesi in FM richiede un minor numero di forme d'onda di partenza per produrre timbri musicalmente più ricchi ed è quindi più diffusa della sintesi additiva.

da un repertorio di forme d'onda possibili, archiviato sotto forma di campioni nella memoria del calcolatore. Fra queste forme d'onda, per esempio, si possono trovare onde sinusoidali, onde quadre e a dente di sega. Le forme d'onda memorizzate aumentano nettamente l'efficienza di un oscillatore, perché il calcolatore deve solo andare a consultare tabelle di campioni della forma d'onda, invece di doverli calcolare. La forma d'onda memorizzata determina anche alcuni aspetti del timbro.

L'oscillatore è controllato da due variabili di ingresso: una determina l'ampiezza del segnale in uscita, l'altra determina la frequenza dei cicli di generazione della forma d'onda. Ambedue gli ingressi possono variare nel tempo, consentendo aumenti o diminuzioni in ampiezza e frequenza. L'ingresso relativo all'ampiezza di un oscillatore per la produzione di suoni spesso è l'uscita di un altro oscillatore che svolge la funzione di generatore di inviluppo per controllare l'attacco e l'estinzione del suono. Come abbiamo già detto, le funzioni di attacco ed estinzione hanno un'influenza importante sul timbro di un suono generato dal calcolatore.

Le altre unità di generazione, sommatore e moltiplicatore, come indicano i nomi stessi, calcolano rispettivamente la somma e il prodotto di due numeri forniti in ingresso. Il sommatore può sommare le uscite di due oscillatori sinusoidali che siano stati «accordati» in modo da dare le parziali dello strumento. Può anche sommare un numero piccolo, variabile con regolarità, al segnale di ingresso che specifica la frequenza di un oscillatore: in questo modo si può simulare un effetto di vibrato. Il moltiplicatore può essere usato in vari modi: come controllo di volume (moltiplicando di un fattore dato l'ingresso che stabilisce il valore di ampiezza di un oscillatore) o come traspositore di altezza (moltiplicando di un fattore dato l'ingresso che stabilisce il valore di frequenza di un oscillatore). Spesso vengono usate anche unità di generazione che producono numeri casuali, per creare sia rumori, sia piccole fluttuazioni casuali di altezza e ampiezza, nei suoni generati, per renderli più vivi e meno meccanici.

All'avvio di un programma di Music V, il compositore «crea» uno strumento selezionando un insieme di unità di generazione e specificando i collegamenti fra tutti i loro ingressi e le loro uscite numeriche. Almeno un'uscita deve andare a un convertitore da digitale ad analogico, che converte i campioni binari in una forma accettabile per pilotare un altoparlante. Il progetto di uno strumento è una funzione altamente creativa, disponibile per un compositore di musica al calcolatore; al contrario, un compositore che si basi su strumenti tradizionali per l'interpretazione delle sue opere può aspettarsi solamente suoni tradizionali. Ma come si può «comporre» un brano musicale usando un programma del genere? Per mezzo di liste di note.

Una lista di note è una istruzione di calcolatore che specifica sostanzialmente le stesse informazioni che vengono date all'esecutore da una nota posta sul pentagramma in una partitura musicale. Specifica quando deve essere eseguita una nota, la sua durata, su quale strumento deve essere eseguita, quali debbono essere la sua altezza e la sua intensità. Spesso vengono incluse anche altre informazioni speciali, relative alla regolazione del timbro dello strumento. Le liste di note non assomigliano alle note che siamo abituati a vedere sul rigo musicale: sono lettere e numeri che il calcolatore interpreta come ingressi per gli «strumenti» che producono campioni, costituiti da unità di generazione.

Per effettuare la sintesi di suoni complessi si possono utilizzare numerose strategie generali. Nei suoi lavori, per esempio, Risset faceva uso di una tecnica chiamata sintesi additiva o somma di parziali. In questa tecnica le singole parziali di un dato suono vengono sintetizzate separatamente (in modo che ciascuna abbia la propria frequenza e il proprio inviluppo indipendentemente da tutte le altre) prima di essere sommate per dare una versione di sintesi del suono. Per mezzo della sintesi additiva, quindi, si possono simulare con buona precisione le parziali non armoniche (parziali le cui frequenze non sono multipli interi della fondamentale) del pianoforte, o le parziali non armoniche, molto evidenti, di campane e tamburi. La sintesi additiva è la tecnica più generale per la sintesi di timbri.

Benché potente, la sintesi additiva è anche costosa e lenta. I timbri possiedono molte parziali e se ciascuna deve essere generata separatamente è necessario effettuare una grande quantità di calcoli. Se poi ciascuna parziale segue un proprio andamento è richiesta anche una grande quantità di informazioni di controllo per generare i loro inviluppi. Di conseguenza molti musicisti hanno cercato scorciatoie che permettessero di generare timbri paragonabili a quelli ottenuti per sintesi additiva, ma con minor fatica. Una delle scorciatoie più degne di nota è la sintesi in modulazione di frequenza, inventata da John M. Chowning della Stanford University: è la tecnica utilizzata nei sintetizzatori digitali oggi più diffusi.

La modulazione di frequenza (in sigla FM, dall'inglese *Frequency Modulation*) di solito è intesa come una tecnica per le radiocomunicazioni, che permette la trasmissione di informazioni mediante la modulazione (cioè la variazione) della frequenza di un segnale ad alta frequenza (la portante) con un segnale d'informazione a bassa frequenza (la modulante). La tecnica di Chowning si basa su portanti e modulanti che hanno frequenze identiche o frequenze dello stesso or-

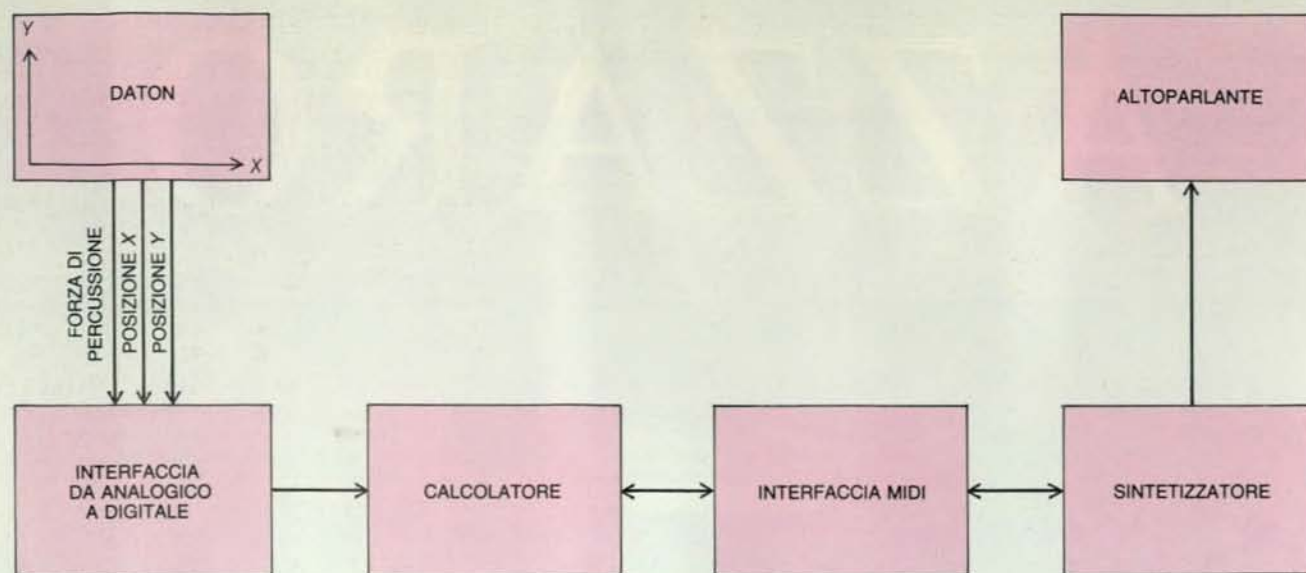
Uno strumento musicale «intelligente» nel laboratorio di uno degli autori (Mathews) è costituito da un calcolatore personale collegato a un sensore appositamente realizzato e a un sintetizzatore digitale di produzione commerciale. Lo strumento permette di «dirigere» un brano di musica programmato: battendo con un martello imbottito sul sensore (che Mathews chiama «daton», contrazione di *drum baton*, ovvero «bacchetta a percussione», e che si può vedere in primo piano), si determina il tempo; variando il punto in cui il martello percuote il sensore si controllano il bilanciamento e il volume dello strumento. Sullo schermo del calcolatore si possono vedere le rappresentazioni di una serie di percussioni: un cerchio il cui raggio è proporzionale alla forza di percussione indica dove il martello ha colpito la superficie del sensore. Al di sopra del calcolatore sono montati il sintetizzatore e un dispositivo di elaborazione digitale dei suoni.

dine di grandezza. Relazioni di frequenza di questo genere fra portante e modulante vengono evitate nelle trasmissioni radio, perché disperderebbero inutilmente il segnale che porta informazioni su una grande larghezza di banda. Nel caso delle forme d'onda musicali questa dispersione della spettro di frequenza del segnale può essere sfruttata utilmente come arricchimento timbrico.

Uno strumento in FM è un po' più complicato da capire di uno strumento ottenuto per sintesi additiva. Nella sintesi in FM sono necessari almeno due oscillatori (uno portante e uno di modulazione). In genere tutti e due producono semplici forme d'onda sinusoidali, il cui andamento nelle fasi di attacco ed estinzione è controllato da generatori di inviluppo. Sostanzialmente la frequenza dell'oscillatore che genera la portante viene variata di una quantità proporzionale all'ampiezza della forma d'onda generata dall'oscillatore modulante. La frequenza della portante, quindi, non è più costante, ma è la somma della frequenza media della portante e dell'uscita, variabile con continuità, dell'oscillatore modulante.

Se la frequenza media della portante e la frequenza modulante sono identiche, il periodo fondamentale dell'onda modulata in frequenza sarà lo stesso della portante non modificata. È cambiata, però, la forma d'onda della portante. Si può dimostrare che, al crescere dell'ampiezza del segnale modulante, aumentano il numero e l'intensità delle parziali armoniche di frequenza più elevata nella forma d'onda portante.

Supponiamo che l'inviluppo della modulante abbia un attacco più morbido di quello della portante. In questo caso le parziali di alta frequenza cresceranno lentamente fino a raggiungere i loro valori di ampiezza nella fase di stato stazionario. È esattamente quello che serve per il timbro di uno strumento della famiglia degli ottoni. In effetti il programma di calcolatore che abbiamo appena descritto produce un timbro da ottone accettabile e richiede solamente due oscillatori e due generatori di inviluppo. A titolo di confronto, per ottenere un timbro analogo con la sintesi additiva sarebbero necessari, tipicamente, 10 oscillatori distinti e 10 distinti generatori di inviluppo.

Se la frequenza della modulante non è uguale a quella della portante, le parziali del segnale modulato in frequenza non sono armoniche. Lo spettro di frequenza del suono è costituito da un grappolo di parziali centrate sulla frequenza della portante e separate l'una dall'altra da un intervallo pari alla frequenza di modulazione. Se la frequenza di modulazione è molto bassa, si forma un insieme di parziali molto denso e il risultato è, in generale, un timbro aspro, dissonante. Se la frequenza della modulante è maggiore di quella della portante, le parziali non armoniche sono molto di-

L'interfaccia digitale per strumenti musicali o MIDI (*Musical-Instrument Digital Interface*) è un protocollo per la codifica digitale di dati musicali; adottato a partire dal 1982 dai costruttori di strumenti elettronici, permette la comunicazione di un sintetizzatore con un calcolatore. Nello strumento di Mathews (*si veda l'illustrazione a pagina 90*) i segnali elettrici analogici, che specificano dove è stata percossa la superficie del daton e quale è stata l'intensità della percussione vengono convertiti in numeri binari e inviati a un calcolatore. Il calcolatore combina quest'informazione con le note archiviate in memoria e invia il tutto, in forma MIDI, al sintetizzatore che esegue le note con i timbri opportuni.

sperse e il risultato è un timbro tipicamente percussivo.

La sintesi in FM non può produrre qualunque spettro di frequenza specificato arbitrariamente, ma Chowning ha dimostrato che si possono produrre efficacemente moltissimi suoni di interesse musicale, semplicemente scegliendo le giuste frequenze e i giusti inviluppi per la portante e la modulante. Inoltre, il suono fondamentale della sintesi FM può essere ulteriormente arricchito semplicemente aggiungendo più forme d'onda ottenute per modulazione di frequenza. Il sintetizzatore Yamaha DX-7, che è basato sulla sintesi in FM, ha sei oscillatori per ciascuna delle 16 voci simultanee. Spesso questi oscillatori sono raggruppati in tre coppie portante-modulante: i segnali in modulazione di frequenza risultanti vengono poi sommati.

La sintesi additiva e la sintesi in FM sono fra i primi metodi di produzione di suoni musicali mediante calcolatore che siano stati inventati e vengono utilizzati ancora oggi. Metodi di sviluppo più recente impiegano dispositivi digitali progettati specificamente a fini musicali. Molti costruttori di sintetizzatori, per esempio, hanno sfruttato la drastica diminuzione di costo degli integrati di memoria per immagazzinare copie di forme d'onda strumentali campionate. Per ogni timbro di strumento, debbono essere memorizzati l'attacco della forma d'onda e una parte del suo stato stazionario o dell'estinzione. Si ottengono poi le note di diversa altezza accelerando o rallentando la velocità a cui sono collegate in successione le forme d'onda memorizzate. La qualità timbrica e la durata possono essere regolate mediando e interpolando parti diverse delle forme d'onda memorizzate.

Le forme d'onda «naturali» memorizzate digitalmente possono produrre suoni «naturali», ma le forme d'onda non possono catturare adeguatamente tutte le sottigliezze di intensità, altezza e timbro nel fraseggio di un passaggio musicale. Per questo si sta lavorando ancora molto per migliorare la «naturalezza» dei suoni di sintesi, compresi quelli costruiti a partire dalle forme d'onda fondamentali, come le onde a dente di sega.

Un'onda a dente di sega è simile alla forma d'onda del suono prodotto dallo sfregamento di una corda di violino, tuttavia quando viene riprodotta attraverso altoparlanti risulta ronzante, poco piacevole. La differenza fra un violino e un altoparlante che riproduca una forma d'onda a dente di sega è che il corpo del violino favorisce naturalmente talune frequenze (le frequenze di risonanza), mentre ne smorza altre. In maniera analoga, il canto della voce umana ha origine da vibrazioni delle corde vocali, che di per sé producono un suono indefinito. Il timbro della voce è formato quasi completamente dalle risonanze del tratto vocale.

Questi principi di acustica musicale furono compresi già agli inizi della musica al calcolatore, ma si sono dimostrati difficili da applicare. Oggi, grazie a nuovi sviluppi (che riguardano principalmente il costo e la velocità dei circuiti integrati) è possibile simulare al calcolatore le risonanze caratteristiche dei sistemi di produzione dei suoni naturali.

La sintesi basata su queste idee ha due aspetti: un'eccitazione generale, come quella fornita da una forma d'onda a dente di sega, e una successiva estinzione graduale di più risonanze principali. La chiave di volta per ottenere dal calcolatore un suono simile a quello del violino, a partire da un'onda a dente di sega, sta nella simulazione dell'estinzione delle oscillazioni di varie componenti di frequenza. Il calcolatore raggiunge questo obiettivo risolvendo equazioni lineari alle differenze finite, il corrispettivo discreto delle equazioni differenziali lineari. La soluzione di un'equazione lineare alle differenze finite è una somma di sinusoidi smorzate, le quali possono simulare le oscillazioni in fase di estinzione di una tavola armonica di violino o del tratto vocale.

Il problema generale che incontra questo tipo di sintesi è la grande quantità di calcoli necessaria per risolvere un'equazione alle differenze finite. Carver A. Mead e John C. Wawrzynek del California Institute of Technology di recente hanno costruito circuiti integrati progettati specificamente per la sintesi musicale mediante soluzione di equazioni lineari alle differenze finite. Se questi chip verranno prodotti commercialmente, la sintesi basata su questa impostazione potrebbe compiere un grande salto in avanti.

Oggi non mancano strategie per la generazione digitale di suoni musicali complessi e soddisfacenti. Nella maggior parte dei casi il problema principale sta nella grande quantità di calcoli necessaria per produrre suoni con un timbro ric-

co. Una stima grossolana ci dà un milione di operazioni (moltiplicazioni e addizioni) per secondo di suono per strumento, con un totale di operazioni variabile dai 10 ai 20 milioni al secondo di suono, nel caso di un lavoro complesso. L'enorme quantità di calcoli necessaria nella sintesi di suoni complessi ha reso impossibile l'effettiva generazione di suoni con calcolatori di uso generale in tempo reale, cioè senza la necessità di registrare i campioni sonori a bassa velocità prima di poterli ascoltare a velocità normale.

I primi calcolatori non avevano nemmeno lontanamente la capacità di sintetizzare musica in tempo reale: i compositori sintetizzavano i loro brani lentamente e pazientemente, registrando i campioni sonori su nastro magnetico digitale, per ascoltare il risultato in riproduzione. In questo modo è stata creata molta musica pregevole, e questo è forse ancora il modo migliore per produrre registrazioni e riempire colonne sonore, perché un compositore può esercitare un grado elevato di controllo nella creazione, nella valutazione e nella revisione delle sue partiture musicali.

Ciononostante, il processo di registrazione elimina un tipo di musicista: l'esecutore «dal vivo». L'interpretazione di sfumature esecutive deve essere scritta nella partitura dal compositore stesso, altrimenti sarà completamente assente. Manca anche il piacere attivo dell'esecuzione musicale, che è molto importante per tutti gli esecutori, professionisti e dilettanti; gli ascoltatori possono anche rimpiangere l'impossibilità di una partecipazione all'esecuzione dei musicisti ai concerti.

Oggi esistono calcolatori con una potenza sufficiente da permettere l'esecuzione musicale in tempo reale, ma in genere sono troppo costosi e ingombranti per assumersi il ruolo di strumenti musicali nelle sale da concerto o nelle case. La soluzione sicuramente sta nella produzione di chip dedicati ed efficienti, come quelli che si trovano nei sintetizzatori Yamaha e quelli prodotti in forma sperimentale da Mead e Wawrzynek. In effetti, gli strumenti musicali digitali basati su questi integrati in molti casi sono meno costosi di alcuni strumenti acustici tradizionali.

Sotto questo profilo, la definizione di un protocollo standard di interfaccia digitale per strumenti musicali (MIDI, *Musical-Instrument Digital Interface*), a cui hanno aderito i costruttori di strumenti elettronici commerciali, è una piacevole novità per il musicista al calcolatore: permette infatti di collegare un calcolatore a questi strumenti, dotandoli di un po' di «intelligenza». Inizialmente l'interfaccia MIDI era stata intesa come uno standard per la trasmissione di informazioni di controllo fra sintetizzatori di marche e modelli diversi. Premendo un tasto sulla tastiera di un sintetizzatore dotato di MIDI, non solo si ottiene l'esecuzione di una nota, ma anche la trasmissione, su un cavo di uscita, di alcuni bit di dati, che dicono quale sia il tasto premuto, e con quale forza è stato abbassato. Un sintetizzatore può avere anche un cavo MIDI di ingresso: se attraverso questo cavo riceve informazioni sul tasto premuto, eseguirà a sua volta una nota esattamente come se fosse stato premuto uno dei suoi tasti. In linea di principio, tutto ciò che può essere fatto su un sintetizzatore può essere controllato localmente da sensori (tasti, pulsanti, manopole o simili) presenti sulla macchina, o controllato a distanza via MIDI.

I creatori dell'interfaccia MIDI forse non avevano in mente di utilizzarla come modo di comunicazione fra un calcolatore e un sintetizzatore, ma essa assolve benissimo questo ruolo. Certo rende piacevole la vita a quanti amano suonare con calcolatori e sintetizzatori: un comune calcolatore personale ha abbastanza potenza per effettuare il controllo musicale, dal momento che la generazione delle forme d'onda è affidata a circuiti specializzati che si trovano all'interno del sintetizzatore stesso. Le parti essenziali di uno strumento intelligente di questo tipo sono il sintetizzatore, un sensore su cui l'esecutore suona, il calcolatore e infine (ma ben lungi dall'essere ultimo in ordine di importanza) il software che coordina le varie componenti in un tutto unico.

Anche se i metodi che abbiamo descritto per la sintesi dei suoni sono stati usati per imitare strumenti tradizionali, le apparecchiature elettroniche digitali possono creare con altrettanta facilità classi completamente nuove di suoni. Cosa ancora più importante, una forma d'onda campionata (non importa se registrata digitalmente da uno strumento «reale» o prodotta direttamente nel calcolatore) si presta agevolmente alla manipolazione. Mediante elaborazione digitale, un suono particolare può essere facilmente trasformato in un suono completamente diverso. Per esempio, gli spettri di frequenza e gli inviluppi che caratterizzano le forme d'onda del parlato di un uomo possono essere plasmati fino a dare, poniamo, il ruggito di un leone.

Studi ulteriori sulla produzione di suoni musicali mediante calcolatore certamente condurranno a imitazioni più precise di strumenti tradizionali, e la definizione di metodi per un controllo fine e veloce delle loro qualità timbriche, controllo che è di grande importanza nelle effettive esecuzioni musicali. I calcolatori, poi, hanno una funzione chiave per poter chiarire le risposte soggettive indotte dai suoni. Questo è un aspetto di particolare importanza per il compositore moderno, che non si limita più ad arrangiare suoni che possono essere prodotti da strumenti convenzionali: ora è possibile evocare qualunque suono immaginabile, e anche qualche suono che immaginabile non è.

(RI)CREAZIONI AL CALCOLATORE

di A. K. Dewdney

Successori in tre dimensioni per il gioco Vita

Vita, il popolare gioco che si basa su un automa cellulare e che si gioca su una griglia bidimensionale, ha ispirato delle versioni tridimensionali. Carter Bays, informatico dell'Università della Carolina del Sud, ha analizzato una grande varietà di versioni tridimensionali e ha individuato le due più promettenti, che chiama Vita 4555 e Vita 5766. Entrambe le versioni riproducono molte caratteristiche del gioco Vita originale (come i lampeggiatori e gli alianti) e una di esse risulterà sicuramente una degna compagna del gioco inventato nel 1968 dal matematico John Horton Conway dell'Università di Cambridge.

Il gioco di Conway, come molti lettori ricorderanno per averne letto più volte nella rubrica «Giochi matematici» di Martin Gardner, si svolge su una griglia bidimensionale infinita di cellule quadrate. Ciascuna cellula ha otto adiacenti (quattro agli angoli e quattro lungo i lati) e può trovarsi nell'uno o nell'altro di due stati, viva o morta. Da qualche parte un grande orologio segna lo scorrere del tempo e a ogni istante certe cellule possono diventare vive e altre morire. Il destino di una cellula è determinato dal numero di adiacenti vivi. Se, per esempio, a un certo istante una cellula viva ha meno di due adiacenti vivi o più di tre, all'istante successivo sarà morta, per la ragione che una cellula viva può essere denutrita o soffocata dalla sovrappopolazione. Una cellula morta, invece, sarà rinata all'istante successivo se ha esattamente tre adiacenti vivi: una nascita richiede tre «genitori».

Conway aveva chiamato Vita il suo gioco perché le cellule potevano essere vive o morte. Presto fu chiaro, però, che il nome era più appropriato di quanto si sospettasse. Varie configurazioni di cellule mostrano un comportamento sorprendentemente complesso e simile a quello degli organismi viventi (*si veda l'illustrazione a pagina 100 in alto*). Il comportamento è ciclico: a ogni battito dell'orologio le configurazioni cambiano, ma dopo un numero finito di istanti compaiono nuovamente le configurazioni originali. Alcune configurazioni rimangono stabili e altre si muovono lungo la griglia spostando una cellula alla volta in direzione orizzontale, verticale o diagonale. Entrambi i tipi di configurazioni hanno nomi di fantasia. Tra quelle stabili ci sono i fari, gli alveari, i lampeggiatori e i blocchi; tra i «viaggiatori» ci sono gli alianti e le astronavi. Vita di Conway, comunque, va molto al di là della semplice imitazione di fenomeni naturali. Come facevo osservare in questa rubrica nel luglio del 1985, è addirittura possibile costruire un calcolatore nel piano cellulare di Vita.

Non dovrebbe sorprendere il fatto che certe versioni del gioco di Conway che si muovono nell'ampio spazio cellulare a tre dimensioni possano dar luogo a fenomeni ancor più fantastici. Queste versioni sono Vita 4555 e Vita 5766 di Carter Bays, in cui ciascuna cellula è un cubo anziché un quadrato e ha 26 adiacenti anziché otto.

I nomi Vita 4555 e Vita 5766 sono tratti da un ristretto lessico elaborato da Bays. I primi due numeri determinano il destino delle cellule vive; il primo numero indica il minimo di adiacenti vivi che una cellula deve avere per non risultare denutrita, mentre il secondo indica il massimo che può averne prima di soffocare per sovrappopolazione. Il terzo e quarto numero regolano il destino delle cellule morte; indicano rispettivamente il minimo e il massimo di adiacenti vivi che una cellula morta deve avere per diventare viva. (In tutte e due le versione di Vita che prenderò in esame in questa sede, il terzo e il quarto numero sono uguali; in generale, però, non è necessariamente così.) Secondo la notazione di Bays, quindi, Vita di Conway diventa Vita 2333.

Vita 4555 è semplice quanto Vita 2333. Una cellula viva muore se ha meno di quattro o più di cinque adiacenti vivi. Una cellula morta diventa viva se ha esattamente cinque adiacenti vivi. Nel corso di una normale ricerca di regole in questo campo di valori, Bays fu attirato per la prima volta da Vita 4555 quando notò una strana configurazione di cellule cubiche che sgusciava dalle profondità dello schermo del suo Macintosh (*si veda l'illustrazione a pagina 100 in basso*). Era un aliante tridimensionale, che si riproduceva dopo essere passato attraverso quattro distinte configurazioni, ciascuna delle quali costituita da 10 cubi in formazione oblunga e stranamente smussata che si muoveva nello spazio come un divano in caduta libera.

Incuriosito, Bays decise di sondare ulteriormente le regole per il 4555, predisponendo un gran numero di sperimentali «brodi promordiali». Seminando in uno spazio iniziale cubi portati casualmente alla vita, mise in moto il suo universo cellulare. In ogni generazione, alcuni cubi morivano e altri diventavano vivi. Il numero di cubi vivi diminuì a ogni generazione, ma non prima che Bays notasse alcuni complessi curiosamente stabili che non cambiavano da una generazione a quella successiva. Alcuni gli ricordavano piedistalli, croci, gradini, palle e bilancieri (*si veda l'illustrazione a pagina 102*). Semine successive portarono ancora ad altre configurazioni stabili e anche a configurazioni cicliche, a cui Bays ha dato nomi bizzarri come rotore e cavallino rampante.

Come un fisico nucleare che abbia nuove particelle con cui giocare, Bays predispose collisioni tra alianti e altre piccole configurazioni con ogni angolazione concepibile. «Tra le collisioni più sorprendenti - rileva - ce n'è una tra un aliante e una palla che dà origine a una massa confusa formata da 29 cubi, il che non è di per sé particolarmente rilevante. Ma improvvisamente ricompare l'aliante, in ritardo di parecchie generazioni e un po' spostato.»

Bays chiama natura i prodotti dei brodi primordiali e anche i risultati delle collisioni degli alianti. Alcune configurazioni sono prodotte molto facilmente; altre, più esotiche, richiedono un lavoro maggiore. Per esempio, si possono agganciare insieme configurazioni a forma di arco per formare una nuova configurazione stabile che Bays chiama porticato. Emerge un'intera architettura di palizzate, scale, pareti e catene. Le pareti possono essere curvate a formare spirali e una grande varietà di forme esotiche stabili limitate solo dall'immaginazione di un entusiasta di Vita 4555 (*si veda l'illustrazione nella pagina a fronte*).

Altri fenomeni richiedono una descrizione in queste pagine. C'è una forma solitaria di sette cubi, chiamata *greeter* («salutatore») che muore a meno che sia in presenza di un altro salutatore. Un aliante che cerchi di attraversare un salutatore viene afferrato e il suo viaggio interrotto. Un secondo aliante può casualmente andare in aiuto del compagno entrando in collisione con il salutatore e facendolo esplodere.

Delle due fruttuose versioni tridimensionali scoperte da Bays, Vita 4555 sembra la sua favorita. È abbastanza curioso che si possano ottenere le cifre di Vita 4555 sommando 2 alle cifre del codice di Vita di Conway, cioè 2333. Forse la coincidenza adombra il fatto che alla fine Vita 4555 si imporrà come degno compagno di Vita di Conway.

Strettamente parlando, tuttavia, Vita 5766 imita il gioco di Conway in modo più rigoroso di quanto faccia Vita 4555. Nella fattispecie, sotto particolari condizioni, Vita 5766 può fornire una simulazione di Vita di Conway sul piano. Le condizioni sono specificate da un teorema trovato da Bays.

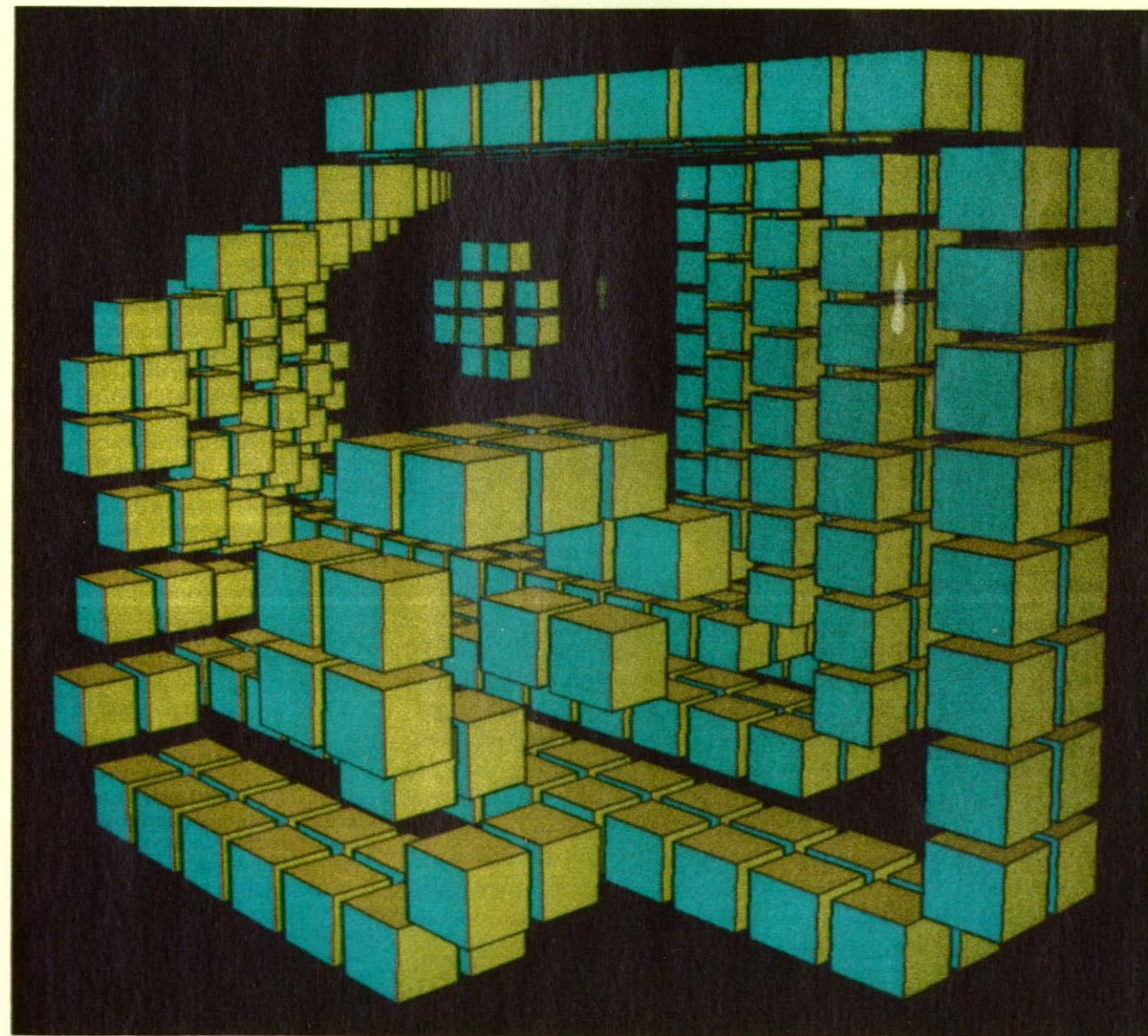
Immaginiamo di guardare la griglia piana di Vita di Conway e di vedervi una particolare configurazione di cellule (quadrate) vive. Poniamo ora un cubo vivo direttamente sopra ciascun quadrato della configurazione e un altro direttamente sotto. Seguendo le regole di Vita 5766, i cubi imiteranno sempre perfettamente il comportamento delle cellule di Conway imprigionate se (e solo se) sono soddisfatte due condizioni:

1. Nessuna cellula quadrata viva del piano ha mai cinque adiacenti vivi.

2. Nessuna cellula quadrata morta del piano ha mai sei adiacenti vivi.

Molte configurazione di Vita, compreso l'aliante di Conway, soddisfano queste condizioni. Si può creare un aliante di Vita 5766 trasformando in cubi un aliante bidimensionale. Esso occupa due strati adiacenti di cellule nello spazio tridimensionale e lì dentro è costretto a muoversi eternamente (*si veda l'illustrazione a pagina 104 in alto*).

Molte altre forme di Vita di Conway, come il faro, l'alveare, l'aliante e il bloc-co, così come configurazioni più esotiche come la barca, l'orologio e l'insegna del barbiere, soddisfano le condizioni del teorema di Bays. Esistono come entità a due strati di Vita 5766 che si comportano esattamente come i loro corrispettivi bidimensionali. Sfortunatamente, non tutte le configurazioni di Vita di Conway soddisfano i criteri. Tra le pecore nere della famiglia vi è il famoso cannone ad alianti scoperto da R. William, Jr., nel 1970 quando studiava al Massachusetts Institute of Technology. Se si esamina il cannone ad alianti a un qualsiasi stadio del suo ciclo vitale, si scoprono cellule vive con cinque adiacenti vivi. Qualsiasi violazione di una delle due condizioni del teorema di Bays ha come conseguenza un distacco dalla simulazione. «Quando questo avviene - afferma Bays - l'oggetto, fino a quel momento confinato in due piani, forma quasi sempre una massa tridimensionale rotondeggiante che di solito muore piuttosto rapidamente.»

Si può fare in modo che Vita 5766 simuli in modo più completo Vita di Conway erigendo quella che Bays chiama

Un aliante si fa strada attraverso una spirale stabile nel gioco Vita 4555 di Carter Bays

una barriera spaziotemporale. Questa barriera è formata da un unico strato di cellule vive con una serie di aperture (*si veda l'illustrazione a pagina 104 in basso*). Le aperture sono disposte in modo che ogni cubo della barriera ha esattamente sette adiacenti. I cubi della barriera evitano di poco la morte per sovrappopolazione, mentre qualsiasi cubo immediatamente adiacente alla barriera avrà più di sei adiacenti vivi e non diventerà mai vivo. La zona morta che si produce da un lato e dall'altro della barriera spaziotemporale può essere sfruttata per riprodurre esattamente Vita di Conway. Si costruisce un sandwich di due barriere spaziotemporali parallele distanti quattro cubi una dall'altra. I due piani centrali, dei quattro che si trovano tra le barriere, possono servire da sostegno per qualsiasi forma di Vita: le due condizioni del teorema di Bays sono soddisfatte da qualsiasi popolazione di cubi vivi che occupi il centro del sandwich.

Nonostante la sua capacità di imitare Vita di Conway in una forma o nell'altra, Vita 5766 manca di quella che si potrebbe chiamare energia cellulare. Secondo Bays, i brodi primordiali casuali di Vita 5766 sembrano sempre «depositarsi» più in fretta di quelli di Vita 4555. La condizione instabile di Vita 4555 suggerisce più ampie possibilità computazionali. In effetti, Vita 4555 abbonda di forme simmetriche stabili e oscillanti. Forse la decisione finale su quale debba essere il miglior compagno tridimensionale per Vita di Conway dipenderà dalla possibilità di far imitare a Vita 4555 il gioco bidimensionale. Se si riuscirà in questo intento, Vita 4555 sarà superiore a Vita 5766 sotto tutti gli aspetti.

Bays ha realizzato un fascicolo di 40 pagine intitolato *The Game of Three-dimensional Life* dinargli presso il Department of Computer Science, University of South Carolina, Columbia, S. C. 29208. Nella monografia c'è tutto quello che ho descritto e molte altre cose, tra cui consigli per un'efficace programmazione dei giochi. Bays dispone anche di un programma per Macintosh che fa funzionare entrambe le forme di Vita tridimensionale.

Almeno in linea di principio, non è difficile scrivere un programma che calcoli e visualizzi successive generazioni dell'una o dell'altra forma di Vita tridimensionale. Si può anche modificare il programma in modo che effettui una ricerca di regole, nella speranza di scoprire una terza forma di Vita tridimensionale che possa essere sfuggita alla ricerca di Bays.

A due grandi matrici tridimensionali, chiamate *cellule* e *nuovecellule*, vengono assegnati tre indici, *i*, *j* e *k*, corrispondenti alle tre coordinate dello spazio cellulare. Il contenuto di ciascun elemento della matrice indica se la cellula corrispondente è viva o morta. Diciamo che 1 sta per la vita e 0 per la morte.

Non sorprende che siano necessari tre cicli annidati per calcolare lo stato di ciascuna cellula per ciascuna generazione. Il ciclo esterno utilizza l'indice *i* per calcolare una successione di piani che attraversa lo spazio. All'interno di questo ciclo ce ne sono altri due che usano, rispettivamente, *j* e *k*. Il ciclo *j* calcola righe successive di ciascun piano e il ciclo *k* calcola cellule successive di una riga. Ci si può servire come guida della seguente forma generica:

```
per i = 1 a 30
  per j = 1 a 30
    per k = 1 a 30
      calcolare adiacenti
      decidere stato
      visualizzare cubi vivi
```

Il numero 30 è naturalmente arbitrario. L'utilizzo, però, di numeri superiori a 30 è consigliabile solo a lettori che abbiano una pazienza infinita, perché il tempo di calcolo diventa incredibilmente lungo.

Nel ciclo più interno, il programma deve affrontare tre compiti fondamentali. Il compito *calcolare adiacenti* richiede che il programma esamini i 26 adiacenti di ciascuna cellula per trovare quanti di essi sono vivi. Questo si può fare con tre piccoli cicli oppure elencando tutte le possibili coordinate delle 26 cellule. In forma ciclica, il procedimento potrebbe usare il seguente algoritmo:

Alcune forme del gioco bidimensionale Vita di John Horton Conway

Un aliante in Vita 4555

```
tot ← 0
per l = i − 1 a i + 1
  per m = j − 1 a j + 1
    per n = k − 1 a k + 1
      se cellule(l,m,n) = 1
        allora tot ← tot + 1
tot ← tot − cellule(i,j,k)
```

L'ultima riga del procedimento garantisce che lo stato di *cellule(i,j,k)* non contribuisca al totale.

Dopo aver stabilito il totale *tot* di adiacenti vivi, il programma deve decidere il nuovo stato della cellula in esame, *cellule(i,j,k)*. Il compito *decidere stato* consiste semplicemente nel controllare la dimensione di *tot* in relazione allo stato di *cellule(i,j,k)*:

```
se cellule(i,j,k) = 0
  allora se tot = 5
    allora nuovecellule(i,j,k,) ← 1
    altrimenti
    nuovecellule(i,j,k) ← 0
se cellule(i,j,k) = 1
  allora se tot < 4 o tot > 5
    allora nuovecellule(i,j,k,) ← 0
    altrimenti
    nuovecellule(i,j,k) ← 1
```

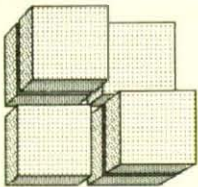
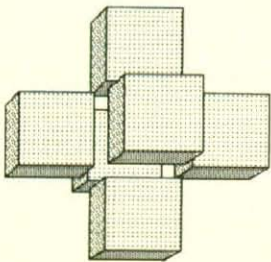
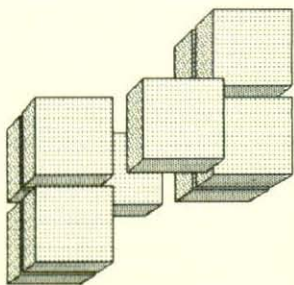
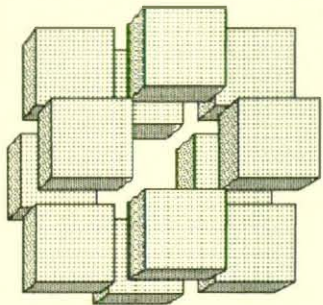
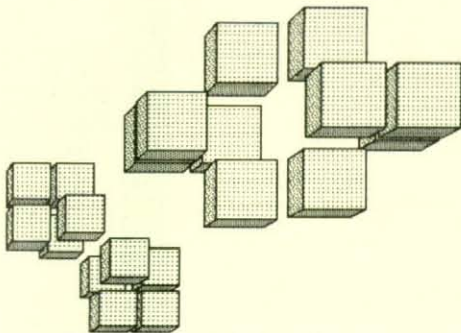
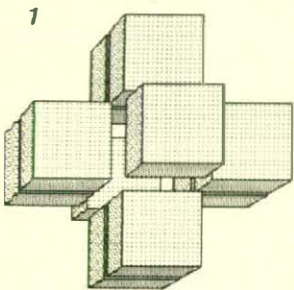
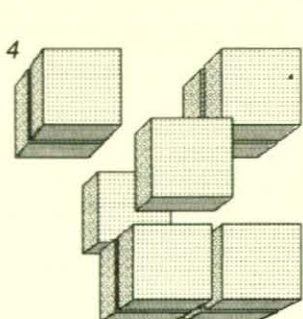
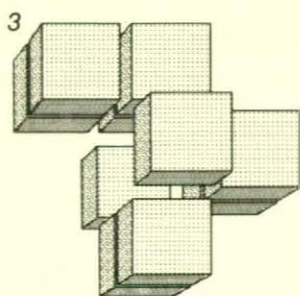
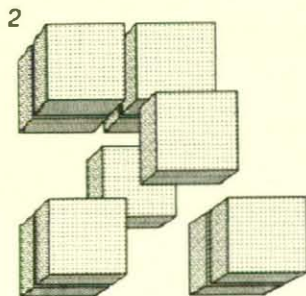
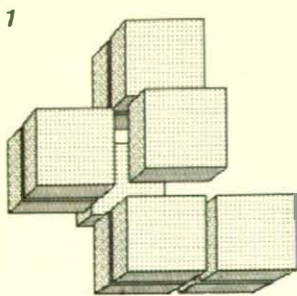
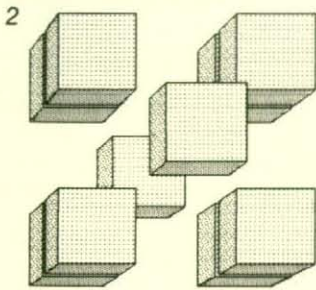
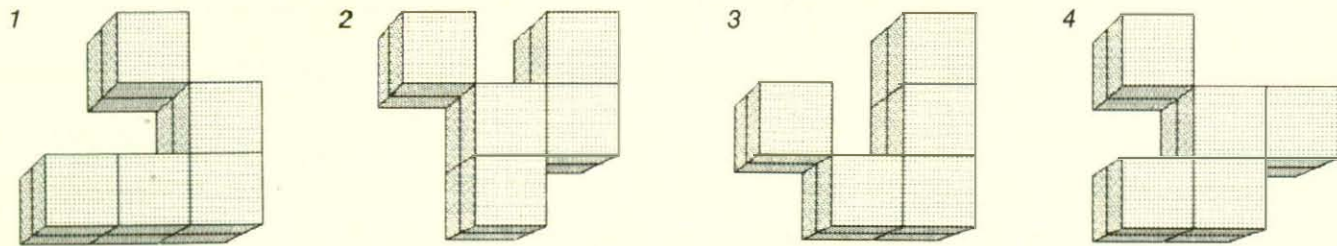
Sono partito dal presupposto che il lettore stia programmando Vita 4555. Volendo si può modificare l'algoritmo in modo che vada bene per Vita 5766 oppure in modo da renderlo abbastanza generale perché possa controllare qualsiasi regola tridimensionale. Mi sembra opportuno fermarci un attimo ad approfondire questo argomento.

Una versione generale del precedente calcolo di stato potrebbe utilizzare quattro variabili chiamate *ai*, *as*, *fi*, *fs*. Le lettere *a* e *f* stanno per ambiente e fecondità, mentre le lettere *i* e *s* stanno per inferiore e superiore. Quindi *ai* e *as* sono i limiti inferiore e superiore perché una cellula possa continuare a vivere nel suo ambiente; la cellula rimarrà viva se il numero di cubi che la circondano è maggiore o uguale ad *ai* ma minore o uguale ad *as*. Allo stesso modo, *fi* e *fs* sono le condizioni di fecondità per una cellula morta, la cui rinascita è garantita se il numero di cellule vive che la circondano è maggiore o uguale a *fi*, ma minore oppure uguale a *fs*. Quindi l'algoritmo generale è:

```
se cellule(i,j,k) = 0
  allora se tot < fi o tot > fs
    allora nuovecellule(i,j,k,) ← 0
    altrimenti
    nuovecellule(i,j,k) ← 1
se cellule(i,j,k) = 1
  allora se tot < ai o tot > as
    allora nuovecellule(i,j,k,) ← 0
    altrimenti
    nuovecellule(i,j,k) ← 1
```

A questo punto, in entrambe le versioni di un programma Vita tridimensionale, il contenuto di *nuovecellule* può essere trasferito in *cellule* per mezzo del triplo

ciclo appropriato, liberando così *nuove-cellule* per la successiva generazione di cubi vivi.

Nello studio finale del processo di calcolo, *visualizzare cubi vivi*, il programma visualizza un certo cubo se esso è vivo. Nel disegnare i cubi, è consigliabile riempire le superfici visibili, perché, se si usano i nudi scheletri, il risultato è una scena confusa e pressoché indecifrabile. Il modo più semplice per assicurarci che i cubi che stanno davanti nascondano correttamente i cubi che stanno dietro è quello di controllare che *i*, l'indice esterno, passi, in relazione all'osservatore, dalla posizione posteriore a quella anteriore dello spazio cellulare. Lo spiacevole fatto che in questo modo si debbano oscurare alcuni cubi mette in luce l'unico svantaggio di qualsiasi forma tridimensionale di Vita: a differenza della versio-

Forme stabili e forme cicliche in Vita 4555

Un aliante nel gioco tridimensionale Vita 5766 di Bays

ne bidimensionale di Conway, non si può spaziare con lo sguardo su tutto quello che avviene. Lo svantaggio di qualsiasi gioco tridimensionale è ugualmente condiviso dal nostro reale mondo tridimensionale. Noi non possiamo vedere tutto quello che succede: e questo è forse una fortuna.

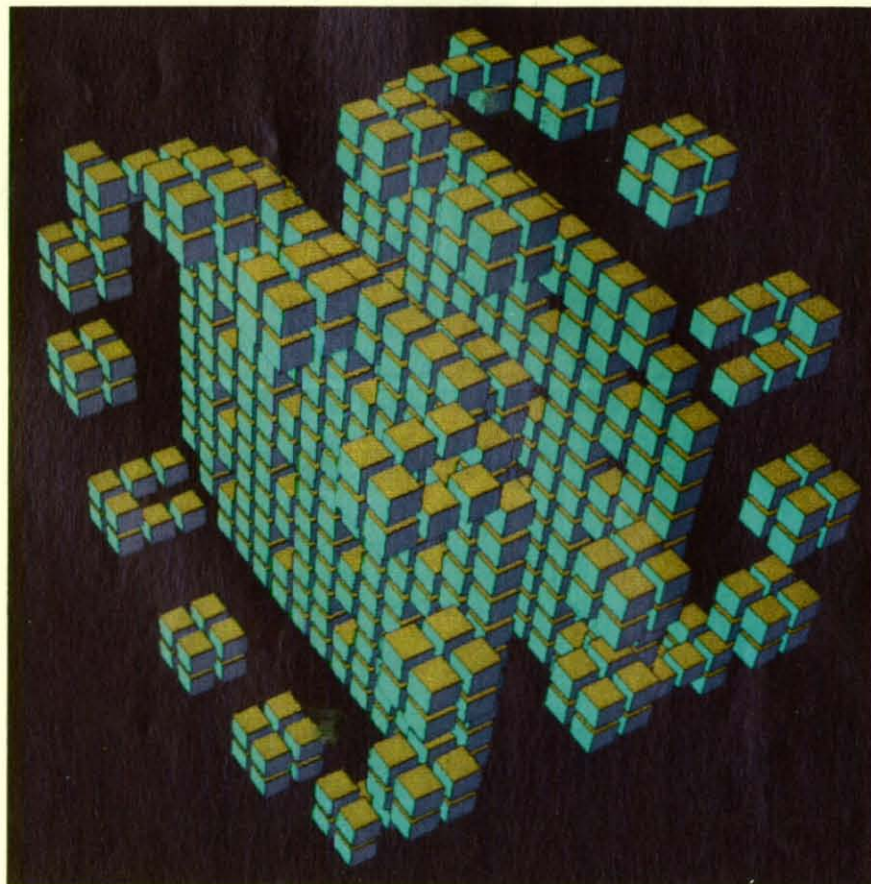
Quando si inserisce lo stadio finale del processo di calcolo, il programma tende a essere piuttosto lento. Una semplificazione che può rendere un po' più veloci le cose consiste nel sostituire i cubi con sfere (in realtà, dischi pieni, di dimensioni variabili a seconda della loro «profondità» nello schermo). Per il resto, la monografia di Bays citata in precedenza contiene molti suggerimenti per migliorare notevolmente la velocità.

Come al solito, sarò lieto di riferire fenomeni inconsueti o interessanti scoperti dai lettori. Prima di tutto, per quegli intrepidi esploratori che richiedono il massimo di generalità, ci sono forme di Vita tridimensionali (degne del titolo) che Bays possa aver trascurato? Poi ci possono essere lettori che desiderano prepararsi da soli i brodi primordiali. Per il resto, ci si può chiedere quali alianti, astronavi, cannoni da alianti e altre configurazioni rimangono da scoprire. Per evitare di riprodurre ciò che Bays già sa, è indispensabile procurarsi il suo *The Game of Three-dimensional Life*.

Il gioco identificato come Star Trek nell'articolo di gennaio esisteva, almeno in ispirito, ben prima della popolare serie televisiva con quel nome. Obert Leonard Nelson, Jr., di Austin, Texas, ricorda che un gioco chiamato Spacewar! (il nonno dei giochi con astronavi che sparano) fu scritto nei primi anni sessanta al Massachusetts Institute of Technology per il calcolatore PDP-1. Secondo un articolo del numero dell'agosto 1981 di «Creative Computing», gli autori del gioco erano professori e studenti del MIT fieri di fregiarsi del nome di «hacker» (termine intraducibile, che indica gli appassionati di calcolatori, *ndr*).

Naturalmente, il PDP-1 era scomodo da programmare e i parametri di ingresso erano determinati con commutatori su una consolle.

Kenneth L. VanEseltine di Schoolcraft, Michigan, ha raffinato il programma di guerra spaziale da me descritto in gennaio. Dapprima lo ha chiamato STAR TRUK («camion delle stelle») a causa della sua lentezza e gli ha dato il nome STAR TREK solo quando ha trovano il modo di eliminare in gran parte dal ciclo principale l'aritmetica in virgola mobile. Ecco alcuni suggerimenti per i lettori. Per evitare il calcolo delle radici quadrate, VanEseltine ha precalcolato una matrice bidimensionale di distanze per ogni possibile punto (x,y) dall'origine $(0,0)$. La distanza tra due punti qualsiasi (x_1, y_1) e (x_2, y_2) viene calcolata consultando nella tabella la distanza per $(x_2 - x_1, y_2 - y_1)$. VanEseltine ha anche tabulato seni e coseni in due matrici di 36 elementi ciascuna, uno per ogni 10 gradi. Ha scoperto anche che i missili sono più facili da controllare se sono sempre 20. In un qualsiasi dato momento, molti possono essere solo virtuali, nel qual caso semplicemente non vengono disegnati. Inoltre, le matrici di velocità non dovrebbero essere bidimensionali, bensì coppie di matrici a una dimensione. Su molti calcolatori, basta questa semplice ridefinizione per ridurre di molto il tempo di elaborazione. Certo i lettori più decisi riscopriranno molti degli altri miglioramenti trovati da VanEseltine.

Infine, per i lettori che vogliano semplicemente provare l'esperienza di essere guerrieri dello spazio, Bill Seiler di Scotts Valley, California, ha un programma chiamato SPACEWAR che gira su PC IBM e su numerose altre macchine. Seiler è lieto di mettere a disposizione su disco il suo programma; scrivetegli, per ulteriori informazioni, al seguente indirizzo: 317 Lockewood Lane, Scotts Valley, California, 95066.

Due barriere spaziotemporali in Vita 5766